Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 130 (2013)

Public vs private coin in bounded-round information

Mark Braverman [*] Ankit Garg [†]

January 15, 2014

Abstract

We precisely characterize the role of private randomness in the ability of Alice to send a message to Bob while minimizing the amount of information revealed to him. We show that if using private randomness a message can be transmitted while revealing I bits of information, the transmission can be simulated without private coins using $I + \log I + O(1)$ bits of information. Moreover, we give an example where this bound is tight: at least $I + \log I - O(1)$ bits are necessary in some cases. Our example also shows that the one-round compression construction of Harsha et al. [HJMR07] cannot be improved.

1 Introduction

In this paper we investigate the role of private randomness in the ability of two parties to communicate while revealing as little information as possible to each other – i.e. to communicate at low information cost. More specifically, Alice and Bob are given possibly correlated inputs X and Y and need to perform a task T by means of a communication protocol π. Alice and Bob share a public random string R; in addition they have access to private random strings R_A and R_B, respectively. The *information cost* of π with respect to a distribution $(X, Y) \sim \mu$ is the quantity

$$\mathsf{IC}_\mu(\pi) := I(\Pi; Y|XRR_A) + I(\Pi; X|YRR_B),$$

where $\Pi = \Pi(X, Y, R, R_A, R_B)$ is the random variable representing the transcript of the protocol.

It is not hard to see that if the goal is to solve a task T while minimizing the information cost of the protocol, we can always avoid using the public randomness string R: to simulate public randomness, before the beginning of the protocol's execution, Alice can send a portion of R_A, which will be used as R for the remainder of the protocol. This modification increases the communication cost of the protocol, but it is not hard to see that it does not change its information cost. Therefore, in the context of information complexity, private randomness is at least as good as public randomness. Is the converse true? In other words, can any protocol π that uses private randomness be simulated by a protocol π' which uses only public randomness so that $\mathsf{IC}_\mu(\pi') \le \mathsf{IC}_\mu(\pi)$? The naïve "solution" to this problem would be to simulate π by using the public randomness to simulate private randomness. The following simple example shows why this approach fails. Consider the protocol π in which $X \in \{0, 1\}^n$. Alice samples a uniformly random

[*] Department of Computer Science, Princeton University, email : mbraverm@cs.princeton.edu. Research supported in part by an Alfred P. Sloan Fellowship, an NSF CAREER award CCF-1149888, and a Turing Centenary Fellowship.

[†] Department of Computer Science, Princeton University, email : garg@cs.princeton.edu.

ISSN 1433-8092

string $R_A \in_U \{0,1\}^n$, and sends the bitwise XOR $M := X \oplus R_A$ to Bob. This protocol conveys 0 information to Bob about X. However, if the public randomness R were to be used to produce R_A, then Bob would also know R_A, and thus the message M reveals $X = M \oplus R_A$ to Bob – drastically increasing the information cost of the protocol. This, of course, does not mean that a more sophisticated simulation scheme cannot work.

It is instructive to compare this question to the public-vs-private randomness question in randomized *communication complexity.* In the context of communication complexity the situation is somewhat reversed: it is obvious that public randomness can be used to simulate private randomness: the parties can always designate part of their public randomness as "private randomness". This will not affect the communication cost of the protocol (although, as seen above, it may affect its information cost). In the reverse direction, Newman [New91] showed that $R_{\epsilon+\delta}(f) \le R_{\epsilon}^{\text{pub}}(f) + O(\log(\frac{n}{\delta}))$. Thus, up to an additive $\log n$, private randomness replaces public randomness in communication complexity. Does a "reverse Newman theorem" hold for information complexity? Can private randomness be replaced with public randomness at a small cost?

This question has been considered by Brody et al. in [BBK$^+$12], which showed a version of the private-by-public simulation for one-round protocols. In the one-round setting, Alice wishes to send Bob her message – a random variable $M = M(X, R_A)$. Obviously, the information cost of this task is just $I(M;X|Y)$. If Bob receives no input, then it is just $I(M;X)$. In this paper we prove tight bounds on the one round private-by-public simulation. Specifically, we show that the cost of simulating a message M of information cost I without the use of private randomness is between I and $I + \log I \pm O(1)$, and that the upper bound is in fact tight in some cases. Previously, [BBK$^+$12] showed a weaker translation to information cost of at most $I + O(\log n)$, where $n = \max(\log|\mathcal{X}|, \log|\mathcal{Y}|)$ – the log of the sizes of the domains of X and Y. Note that it is always the case that $I \le H(X) \le \log|\mathcal{X}| \le n$, and therefore $\log I \le \log n$. Our lower bound example shows that even if dependence on n is allowed, one cannot do with less than $\log n$ additive overhead.

It is interesting to consider the connection between the problem of simulating a protocol without private randomness, and the problem of compressing communication protocols. The general protocol compression problem [BBCR10, Bra12] is the problem of simulating a protocol π with communication cost C and information cost I with a protocol π' of communication cost C' that is as close to I as possible. The problem of compressing interactive communication is essentially equivalent to the direct sum problem for randomized communication complexity [BR11]. The best known general compression results gives $C' = \tilde{O}(\sqrt{I \cdot C})$, and it is wide open whether $C' = O(I \cdot (\log C)^{O(1)})$ is possible. It has been shown in [BBK$^+$12] (and independently in [Pan12]) that if a protocol π does not use private randomness, then it can be compressed to $O(I \cdot (\log C)^{O(1)})$. Thus a way to replace private randomness with public randomness for unbounded-round protocols would imply a substantial improvement in the state-of-the-art on protocol compression.

Another interesting connection between removing private randomness and compression is in the context of one-message protocols. In the setting where Bob has no input Y, the information cost of sending a message M is just $I := I(M;X)$. Harsha et al. [HJMR07] showed how to simulate such a transmission using $I + O(\log I)$ bits of (expected) *communication* (with access to public randomness). Their work left open the interesting question of whether the additive $O(\log I)$ is necessary. As noted above, a communication protocol with communication C can always be simulated by a protocol with same communication and only public randomness. As information cost is bounded from above by communication cost, a compression scheme is in particular a private-by-public scheme. Thus our lower bound gives an example showing that the $O(\log I)$ additive overhead in [HJMR07] is necessary.

Results and techniques

Our main result gives an upper and lower bound on simulating private randomness by public randomness for one-message protocols.

Theorem 1.1. *Let X, Y be inputs to Alice and Bob respectively distributed according to a distribution μ. Alice and Bob have access to public randomness R', and Alice has access to private randomness R_A. Let π be a protocol where Alice sends a message $M = M(X, R', R_A)$ to Bob, so that the information cost of π is $I := I(X; M|YR')$. Then*

1. *π can be simulated by a one-message public-coin protocol π' such that $\mathsf{IC}_\mu(\pi') \leq I + \log I + O(1)$.*
2. *for each I, there is an example with no Y (i.e. Bob has no "private" knowledge), and no R', such that if $I := I(X; M)$, then any public-coin protocol π' simulating the transmission of M must have information cost of at least $I + \log I - O(1)$.*

Thus, up to an additive constant, our bounds are tight. Note that while the upper bound holds under the most general conditions, for the lower bound it is sufficient to consider protocols without Y (this is the type of protocols considered, for example, in [HJMR07]).

Both the upper and lower bound require some careful analysis. For the upper bound, a natural variant of the one-round compression scheme of Braverman and Rao [BR11] is used. The main challenge is in analyzing the information cost of the resulting public randomness protocol: we need to prove that Bob does not learn too much about X from Alice's message. Suppose that given X and the public randomness R of the simulating protocol, Alice's message in the simulating protocol is $S = S(X, R)$. Observe that in this case

$$I(S; X|YR) = H(S|YR) - H(S|XYR) = H(S|YR).$$

To establish an upper bound on $H(S|YR)$, we show how, someone knowing X, Y and R, can describe S to Bob using a message M' (i.e. $H(S|M'YR) = 0$) such that

$$H(M') \leq I + \log(I) + O(1)$$

Noting that this expression is an upper bound for $H(S|YR)$, completes the proof.

To prove the lower bound, we give a family of specific examples whose information cost necessarily increases by $\log I - O(1)$ when private randomness is replaced with public randomness. Details of the construction are given in Section 4, here we only give the high level idea for why the information cost increases in lieu of private randomness. Consider the following example: Alice knows a secret random string $PASS$ of 128 bits (which we can think of as her password). She wants to send Bob a message M such that $M = PASS$ with probability 1/2 and $M = RANDOM$ with probability 1/2 – that is, half of the time she sends her password and half the time she sends a random 128-bit string. The message M reveals approximately 63 bits of information about $PASS$. To see this, note that given M the posterior distribution of $PASS$ puts mass 1/2 on M and mass 1/2 on the remaining $2^{128} - 1$ strings. The entropy of this distribution is $\approx \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 129 = 65$, down from the prior entropy of 128. Thus $I(M; PASS) \approx 128 - 65 = 63$ bits. One might have expected this number to be 64 bits. Indeed, if Alice had told Bob which of the two cases has occurred, M would reveal $\frac{1}{2} \cdot 128 + \frac{1}{2} \cdot 0 = 64$ bits of information. However, not knowing whether Alice's message is the password or a random string "saves" one bit in information cost. Now suppose Alice was not allowed to use private randomness. Then, intuitively, the public random string R should reveal to Bob whether $M = PASS$ or $M = RANDOM$. Therefore, the information cost of a public-randomness protocol increases to 64 bits. Generalizing from this example, we construct a situation where Alice sends a binary message M of length n and information cost $I \approx n/2 - \log n$, so that any public randomness simulation of M requires information cost of $\geq n/2 - O(1) = I + \log I - O(1)$ – demonstrating the desired gap.

Let us have a look at another example. Suppose that Alice gets a bit $X \sim B_{\frac{1}{2}}$ and she wants to transmit this bit to Bob with error $\frac{1}{2} - \epsilon$. Consider a private-coin protocol in which Alice samples a $B_{\frac{1}{2}+\epsilon}$ bit R. She sends X if $R = 1$ and a $\neg X$ if $R = 0$. Clearly the protocol performs the task of transmitting the bit with error $\frac{1}{2} - \epsilon$. Let Π denote the random variable for Alice's message. The information cost of this protocol is

$$I(\Pi; X) = \frac{1}{2} D(\Pi_0 || \Pi) + \frac{1}{2} D(\Pi_1 || \Pi) = \frac{1}{2} D(1/2 - \epsilon || 1/2) + \frac{1}{2} D(1/2 + \epsilon || 1/2) = \frac{2}{\ln 2} \epsilon^2 \pm o(\epsilon^2)$$

However if we don't allow private coins, then the information complexity of this task is $\geq 2\epsilon$. To see this consider a public-coin protocol that transmits X with error probability $\leq \frac{1}{2} - \epsilon$. It is basically a function $f : \{0,1\} \times \mathcal{R} \to \{0,1\}$ (in case Alice sends a longer message and then Bob applies a deterministic function to that, f could be the composition of those two functions) such that $\mathbb{E}_{r \sim \mathcal{R}}[f(0,r)] = \frac{1}{2} - \epsilon$ and $\mathbb{E}_{r \sim \mathcal{R}}[f(1,r)] = \frac{1}{2} + \epsilon$. Then $Pr_{r \sim \mathcal{R}}[f(1,r) = 1, f(0,r) = 0] \geq 2\epsilon$. Hence

$$I(f(X,R); X|R) = H(f(X,R)|R) = \mathbb{E}_{r \sim \mathcal{R}} H(f(X,r)) \geq 2\epsilon$$

since if $f(1,r) = 1, f(0,r) = 0$, then $H(f(X,r)) = 1$. This example, in some sense, highlights the information-cost advantage one gains from having access to private randomness. It will be interesting to see if this advantage can be amplified over multiple rounds to get a separation between unbounded round private-coin information complexity and public-coin information complexity, and thus in particular between information complexity and communication complexity.

Open problems

Our lower bound example is really about the simulation of a protocol and not about solving a boolean function. So it will be nice to get a 1-round gap for a boolean function. Also it would be nice to get a bigger separation between r-round public-coin information complexity and private-coin information complexity, where r is a constant. Note that using the 1-round example, we can also construct a 2-round example by requiring both Alice and Bob to perform the 1-round task.

1. Does there exist a boolean function f for which 0-error private-coin information complexity is I but 0-error public-coin information complexity is $\geq I + \log(I) - O(1)$?

2. Does there exist a (family of) 3-round private-coin protocol(s) π such that information cost of π is I but any 3-round public-coin protocol simulating π has information cost $\geq I + 3\log(I) - O(1)$?

Acknowledgments

We would like to thank Omri Weinstein for the helpful comments on earlier versions of this paper. We would like to thank Anup Rao and Makrand Sinha for pointing out a bug in the previous version of this paper.

2 Preliminaries

2.1 Communication Complexity

In the two-party communication model, the parties, traditionally called Alice and Bob, are trying to collaboratively compute a known Boolean function $f : \mathcal{X} \times \mathcal{Y}$. Each party is computationally unbounded; however, Alice is only given input $x \in \mathcal{X}$ and Bob is only given $y \in \mathcal{Y}$. In order to compute $f(x, y)$, Alice and Bob communicate in accordance with an agreed-upon communication protocol π. Protocol π specifies as a function of transmitted bits only whether the communication is over and, if not, who sends the next bit. Moreover, π specifies as a function of the transmitted bits and x the value of the next bit to be sent by Alice. Similarly for Bob. The communication is over when *both parties* know the value of $f(x, y)$. The cost of the protocol π is the number of bits exchanged on the worst input. *The transcript* of a protocol is a concatenation of all the bits exchanged during the execution of the protocol.

There are several ways in which the deterministic communication model can be extended to include randomness. In the *public-coin model*, Alice and Bob have access to a shared random string r chosen according to some probability distribution. The only difference in the definition of a protocol is that now the protocol π specifies the next bit to be sent by Alice as a function of x, the already transmitted bits, and a random string r. Similarly for Bob. This process can also be viewed as the two players having an agreed-upon distribution on deterministic protocols. Then the players jointly sample a protocol from this distribution. In the *private-coin model*, Alice has access to a random string r_A hidden from Bob, and Bob has access to a random string r_B hidden from Alice.

Definition 2.1 (Randomized Communication Complexity)**.** For a function $f : \mathcal{X} \times \mathcal{Y} \to Z$ and a parameter $\epsilon > 0$, $R_\epsilon(f)$ denotes the communication cost of the best randomized private-coin protocol for computing f with error at most ϵ on *every* input. Similarly $R_\epsilon^{\text{pub}}(f)$ denotes the cost of the best randomized public-coin protocol for computing f with error at most ϵ on *every* input.

Definition 2.2. We will say that a (randomized) protocol ϕ *simulates* a protocol π if there is a deterministic function g such that $g(\Phi(x, y, R^\phi, R_A^\phi, R_B^\phi))$ is equal in distribution to $\Pi(x, y, R^\pi, R_A^\pi, R_B^\pi)$, $\forall x, y$. Here $R^\phi, R_A^\phi, R_B^\phi$ are the public and private randomness of protocol ϕ and Φ is the random variable for the transcript. Similarly for π.

For the pre-1997 results on communication complexity, see the excellent book by Kushilevitz and Nisan [KN97].

2.2 Information Theory

In this section we briefly provide the essential information-theoretic concepts required to understand the rest of the paper. For a thorough introduction to the area of information theory, the reader should consult the classical book by Cover and Thomas [CT91]. Unless stated otherwise, all log's in this paper are base-2.

Definition 2.3. Let μ be a probability distribution on sample space Ω. *Shannon entropy* (or just *entropy*) of μ, denoted by $H(\mu)$, is defined as $H(\mu) := \sum_{\omega\in\Omega} \mu(\omega) \log \frac{1}{\mu(\omega)}$.

For a random variable A we shall write $H(A)$ to denote the entropy of the induced distribution on the range of A. The same also holds for other information-theoretic quantities appearing later in this section.

Definition 2.4. *Conditional entropy* of a random variable A conditioned on B is defined as

$$H(A|B) = \mathbb{E}_b(H(A|B = b)).$$

Fact 2.5. $H(AB) = H(A) + H(B|A)$.

Definition 2.6. The *mutual information* between two random variable A and B, denoted by $I(A;B)$ is defined as

$$I(A;B) := H(A) - H(A|B) = H(B) - H(B|A).$$

The *conditional mutual information* between A and B given C, denoted by $I(A;B|C)$, is defined as

$$I(A;B|C) := H(A|C) - H(A|BC) = H(B|C) - H(B|AC).$$

Fact 2.7 (Chain Rule)**.** *Let A_1, A_2, B, C be random variables. Then*

$$I(A_1A_2;B|C) = I(A_1;B|C) + I(A_2;B|A_1C).$$

Definition 2.8. Given two probability distributions μ_1 and μ_2 on the same sample space Ω such that $(\forall\omega \in \Omega)(\mu_2(\omega) = 0 \Rightarrow \mu_1(\omega) = 0)$, the *Kullback-Leibler Divergence* between is defined as

$$D(\mu_1||\mu_2) = \sum_{\omega\in\Omega} \mu_1(\omega) \log \frac{\mu_1(\omega)}{\mu_2(\omega)}.$$

For Bernoulli distributions B_p and B_q, we will slightly abuse notation and denote $D(B_p||B_q)$ by $D(p||q)$. The connection between the mutual information and the Kullback-Leibler divergence is provided by the following fact.

Fact 2.9. *For random variables A, B, and C we have*

$$I(A;B|C) = \mathbb{E}_{b,c}(D(A_{bc}||A_c)).$$

Fact 2.10. *Let A, B, C, D, E be random variables.If C, D determine E and $D \to CE \to AB$ is Markov chain, then*

$$I(A;B|CE) = I(A;B|CD)$$

Proof. $I(A;B|CD) = I(A;B|CDE)$, since C, D determine E. Now consider $I(A;BD|CE)$

$$I(A;BD|CE) = I(A;B|CE) + I(A;D|BCE) = I(A;B|CE)$$

Also

$$I(A;BD|CE) = I(A;D|CE) + I(A;B|CDE) = I(A;B|CDE)$$

which completes the proof. □

2.3 Information Complexity

A much more detailed discussion of information complexity and its applications can be found in [CSWY01, BYJKS04, BBCR10, BGPW13] and references therein.

Definition 2.11. The *internal information cost* of a protocol π with respect to a distribution μ on inputs from $\mathcal{X} \times \mathcal{Y}$ is defined as

$$\mathrm{IC}_\mu(\pi) := I(\Pi; X|YRR_B) + I(\Pi; Y|XRR_A).$$

where $\Pi = \Pi(X, Y, R, R_A, R_B)$ is the random variable denoting the transcript of the protocol, R is the public randomness and R_A and R_B are the private random strings of Alice and Bob, respectively. In the previous works, it is defined in a different way (without the conditioning on private random strings), but both the definitions are in fact equivalent.

The following simple fact asserts that information cost is bounded by the communication cost of the protocol (see, e.g. [BR11]):

Lemma 2.12. *For any distribution μ, $\mathsf{IC}_\mu(\pi) \leq CC(\pi)$.*

The *information complexity* of f with respect to μ is

$$\mathrm{IC}_\mu(f, \epsilon) := \inf_\pi \mathrm{IC}_\mu(\pi),$$

where the infimum ranges over all (randomized) protocols π solving f with error at most ϵ when inputs are sampled according to μ.

3 Upper Bound

Theorem 3.1. *Let X, Y be inputs to Alice and Bob respectively distributed according to a distribution μ. Alice and Bob have access to public randomness R', and Alice has access to private randomness R_A. Let π be a protocol where Alice sends a message $M = M(X, R', R_A)$ to Bob, so that the information cost of π is $I := I(X; M|YR')$. Then π can be simulated by a one-message public-coin protocol π' such that $\mathsf{IC}_\mu(\pi') \leq I + \log I + O(1)$.*

Proof. We can assume wlog that R' is a part of M, since $I(X; M|YR') = I(X; MR'|Y)$. Let $\mathcal{U}$ be the message space of the message M. Consider the protocol π' defined in Figure 1.

1. Using public randomness, Alice and Bob get samples $\{(u_i, p_i)\}_{i \geq 1}$, where (u_i, p_i) uniformly sampled from $\mathcal{U} \times [0, 1]$.
2. Let P denote the distribution $M_x = M|_{X=x}$ and Q denote the distribution $M_y = M|_{Y=y}$. Alice sends Bob the index of the first sample, s, such that $p_s < P(u_s)$. Bob decodes this message as being u_s

Protocol 1: Protocol π'

It is clear that Bob's decoding of the Alice's message on input x is distributed according to $P = M_x$. What remains is to analyze the information cost of the protocol. Let R denote the random variable for the public randomness and let S denote the random variable for Alice's message (the index). Then

$$I(S;X|YR) = H(S|YR) - H(S|XYR) = H(S|YR)$$

because S is determined by X and R. It seems difficult to get a handle on $H(S|YR)$, but we can use the following trick : If someone (who knows X, Y, R) can describe to Bob S using a message M' (i.e. S is fixed given M', Y and R), then $H(S|YR) \leq H(M')$. This is because :

$$H(S|YR) + H(M'|SYR) = H(M'S|YR) = H(M'|YR) + H(S|M'YR) = H(M'|YR) \leq H(M').$$

Note that since Alice doesn't know Y, she won't be able to compute M' and hence it does not seem possible for Alice to send the message M using a low communication protocol. To achieve low communication, interaction seems necessary, and this problem has been well studied in [BR11] and [BRWY13]. Now let us describe the message M'. Let P denote the distribution M_x and Q denote the distribution M_y. The message will consist of three parts. The first part would be $k = \lceil \frac{S}{|\mathcal{U}|} \rceil$. The second part would consist of the ceiling of the Q-height of the S^{th} sample i.e. $t = \lceil \frac{p_S}{Q(u_S)} \rceil$. The third part would consist of the index l of the sample Alice wants to send among indices $\{(k-1) \cdot |\mathcal{U}| + 1, \ldots, k \cdot |\mathcal{U}|\}$ that have Q-height between $t-1$ and t.

Now let us look at $\mathbb{E}[|M'||X = x, Y = y]$. We'll analyze the lengths of the three different parts of M' separately.

1. For (u, p) randomly sampled from $\mathcal{U} \times [0,1]$,

   $$Pr[p < P(u)] = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} P(u) = \frac{1}{|\mathcal{U}|}$$

   Thus $Pr[S > r \cdot |\mathcal{U}|] = \left(1 - \frac{1}{|\mathcal{U}|}\right)^{r \cdot |\mathcal{U}|} \leq e^{-r}$. Thus $Pr[k > r] \leq e^{-r}$. Thus

   $$\mathbb{E}[k] = \sum_{r=0}^{\infty} Pr[k > r] \leq 1 + \frac{1}{e} + \frac{1}{e^2} + \ldots = O(1)$$

   Hence $\mathbb{E}[\lceil \log(k) \rceil] = O(1)$.

2. For the S^{th} sample, $p_S < P(u_S)$. Thus $\mathbb{E}[\lceil \log(t) \rceil] \leq \mathbb{E}\left[\log\left(\frac{p}{Q(u)} + 1\right) | p < P(u)\right]$. Since $\log(x+1) - \log(x) \leq \frac{\log(e)}{x}$ (by Lagrange's Mean Value Theorem),

   $$\log\left(\frac{p}{Q(u)} + 1\right) \leq \log\left(\frac{p}{Q(u)}\right) + O\left(\frac{Q(u)}{p}\right)$$

   Thus

   $$\begin{aligned}
   \mathbb{E}\left[\log\left(\frac{p}{Q(u)} + 1\right) | p < P(u)\right] &= \sum_{u \in \mathcal{U}} P(u) \cdot \left(\frac{1}{P(u)} \int_0^{P(u)} \log\left(\frac{p}{Q(u)} + 1\right) \mathrm{d}u\right) \\
   &\leq \sum_{u \in \mathcal{U}} P(u) \cdot \left(\frac{1}{P(u)} \int_0^{P(u)} \log\left(\frac{P(u)}{Q(u)} + 1\right) \mathrm{d}u\right) \\
   &\leq \sum_{u \in \mathcal{U}} P(u) \cdot \left(\frac{1}{P(u)} \int_0^{P(u)} \log\left(\frac{P(u)}{Q(u)}\right) + O\left(\frac{Q(u)}{P(u)}\right) \mathrm{d}u\right) \\
   &= D(P||Q) + O(1)
   \end{aligned}$$

Hence $\mathbb{E}[\lceil \log(t) \rceil] \leq D(P||Q) + O(1)$.

3. For (u, p) randomly sampled from $\mathcal{U} \times [0,1]$,

$$\begin{aligned} Pr[(t-1) \cdot Q(u) < p \leq t \cdot Q(u) | p > P(u)] &\leq Pr[(t-1) \cdot Q(u) < p \leq t \cdot Q(u)]/Pr[p > P(u)] \\ &= Pr[(t-1) \cdot Q(u) < p \leq t \cdot Q(u)]/(1 - \frac{1}{|\mathcal{U}|}) \\ &\leq 2Pr[(t-1) \cdot Q(u) < p \leq t \cdot Q(u)] \\ &= \frac{2}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} Q(u) = \frac{2}{|\mathcal{U}|} \end{aligned}$$

Thus among the indices $\{(k-1) \cdot |\mathcal{U}| + 1, \ldots, k \cdot |\mathcal{U}|\}$, in expectation, there are a constant number that have Q-height between $t-1$ and t. Thus $\mathbb{E}[\lceil \log(l) \rceil] = O(1)$.

Note that for the darts appearing before the dart S, the probability of appearing in some region increases slightly, since they are conditioned on not falling under the histogram of P but the probability increases at most by a factor of 2.

Hence $\mathbb{E}[|M'||X = x, Y = y] \leq D(M_x||M_y) + O(1)$. Now

$$\mathbb{E}[|M'|] = \mathbb{E}_{x,y}\left[\mathbb{E}[|M'||X = x, Y = y]\right] \leq \mathbb{E}_{x,y}[D(M_x||M_y)] + O(1) = I(M; X|Y) + O(1)$$

Now we will use the following lemma to bound $H(M')$.

Lemma 3.2. *Let P be a distribution on the natural numbers such that $\sum_{n \geq 1} P_n \cdot \lceil \log(n) \rceil = I$. Then $H(P) \leq I + \log(I) + O(1)$.*

The lemma says that if the expected length of the numbers is bounded by I, then the entropy is bounded by $I + \log(I) + O(1)$. A bound of $I + 2\log(I) + O(1)$ or of $I + \log(I) + 2\log(\log(I)) + O(1)$ is easy to get via prefix-free encoding of integers, but the fact that we can bound the entropy by $I + \log(I) + O(1)$ is somewhat surprising.

Using the lemma, we get that $H(M') \leq I(M; X|Y) + \log(I(M; X|Y)) + O(1)$, and thus $I(S; X|YR) \leq I(M; X|Y) + \log(I(M; X|Y)) + O(1)$. It remains to prove the lemma.

Proof. (Of Lemma 3.2) Let p_i be the probability mass on the integers between 2^{i-1} and 2^i i.e. $p_i = \sum_{n=2^{i-1}+1}^{2^i} P_n$. Then $\sum_{i=1}^{\infty} i \cdot p_i = \sum_{n \geq 1} P_n \cdot \lceil \log(n) \rceil = I$.

$$H(P) = \sum_{n \geq 1} P_n \log\left(\frac{1}{P_n}\right) \leq P_1 \log\left(\frac{1}{P_1}\right) + \sum_{i \geq 1} p_i \log\left(\frac{2^{i-1}}{p_i}\right) = I \pm O(1) + H(p)$$

The inequality follows from log-sum inequality,

$$\sum_k a_k \log\left(\frac{a_k}{b_k}\right) \geq \left(\sum_k a_k\right) \log\left(\frac{\sum_k a_k}{\sum_k b_k}\right)$$

Then, $\sum_{n=2^{i-1}+1}^{2^i} P_n \log(P_n) \geq p_i \log(p_i/2^{i-1})$. Now let q_j be the probability mass of p_i from $2^{j-1}+1$ to 2^j i.e. $q_j = \sum_{i=2^{j-1}+1}^{2^j} p_i$. Then $\sum_{i \geq 1} i \cdot p_i \geq \sum_{j \geq 1} 2^{j-1} \cdot q_j$. Thus $\sum_{j \geq 1} 2^j \cdot q_j \leq 2I$. Again by the log-sum inequality,

$$H(p) \leq p_1 \log\left(\frac{1}{p_1}\right) + \sum_{j \geq 1} q_j \log\left(\frac{2^{j-1}}{q_j}\right) + O(1) = \sum_{j \geq 1} j \cdot q_j + H(q) \pm O(1)$$

We can assume wlog that I is a power of 2. If $j = \log(2I) + k$, for $k \geq 2$, then $q_j \leq \frac{1}{2^k}$ and hence $q_j \log\left(\frac{1}{q_j}\right) \leq \frac{k}{2^k}$, since $q \log\left(\frac{1}{q}\right)$ is increasing in the interval $(0, \frac{1}{e}]$. Thus $\sum_{j>\log(2I)} q_j \log\left(\frac{1}{q_j}\right) = O(1)$. Let $q = \sum_{j>\log(2I)} q_j$. Since $q_{\log(2I)+k} \leq \frac{1}{2^k}$, $\sum_{j>\log(2I)} j \cdot q_j \leq q \cdot \log(2I) + O(1)$. So all that is needed is to prove that

$$\sum_{j \leq \log(2I)} j \cdot q_j + \sum_{j \leq \log(2I)} q_j \log\left(\frac{1}{q_j}\right) \leq (1-q) \cdot \log(2I) + O(1)$$

Let us look at $j \cdot q_j + \log(2I) \cdot q_{\log(2I)} + q_j \log\left(\frac{1}{q_j}\right) + q_{\log(2I)} \log\left(\frac{1}{q_{\log(2I)}}\right)$. If we decrease q_j and increase $q_{\log(2I)}$ by the same amount, the rate at which $j \cdot q_j + \log(2I) \cdot q_{\log(2I)}$ increases is $\log(2I) - j$. Also $\left(q \log\left(\frac{1}{q}\right)\right)' = \log(e) \cdot \left(\ln\left(\frac{1}{q}\right) - 1\right)$. The difference in rates for $q_{\log(2I)}$ and q_j is $\log\left(\frac{1}{q_{\log(2I)}}\right) - \log\left(\frac{1}{q_j}\right)$. So as long as

$$\log\left(\frac{1}{q_j}\right) - \log\left(\frac{1}{q_{\log(2I)}}\right) \leq \log(2I) - j$$

increasing $q_{\log(2I)}$ and decreasing q_j (by the same amount) will increase $\sum_{j \leq \log(2I)} j \cdot q_j + \sum_{j \leq \log(2I)} q_j \log\left(\frac{1}{q_j}\right)$. Thus we can assume wlog that, $q_j \leq \frac{q_{\log(2I)}}{2^{\log(2I)-j}}$. Now for these values of q_j, it is easy to check that $\sum_{j \leq \log(2I)} q_j \log\left(\frac{1}{q_j}\right) = O(1)$. Also $\sum_{j \leq \log(2I)} j \cdot q_j \leq (1-q) \cdot \log(2I)$ is trivially true. This completes the proof. Note that it is not always true that $\sum_{j \leq \log(2I)} q_j \log\left(\frac{1}{q_j}\right) = O(1)$ but for the distribution maximizing $\sum_{j \leq \log(2I)} j \cdot q_j + \sum_{j \leq \log(2I)} q_j \log\left(\frac{1}{q_j}\right)$, this is true. □

□

We mention a few easy corollaries :

Corollary 3.3. *Let X, Y be inputs to Alice and Bob respectively distributed according to a distribution μ. Suppose that π is a private-coin r-round protocol with information cost $IC_\mu(\pi) = I$. Then π can be simulated by a r-round public-coin protocol π' with information cost $IC_\mu(\pi') \leq I + r\log(I/r) + O(r)$.*

Proof is in the appendix.

Our upper bound also improves slightly the bound of Harsha et al. [HJMR07]. In their setting, Alice wants to send a message M with $I(M; X) = I$ to Bob using low communication and public-randomness is allowed. They give protocol with communication cost $I + \log(I) + \log(\log(I)) + \ldots$. We can get a bound of $I + \log(I) + O(1)$ which is tight (even in terms of public-coin information) as shown by the lower bound in next section. The savings essentially come from the surprising Lemma 3.2.

Corollary 3.4. *Suppose Alice wants to help Bob to sample from the distribution $M|X = x$ and they have access to shared randomness. Let $I(M; X) = I$. Then there exists a public-coin protocol π with expected communication $\leq I + \log(I) + O(1)$, which achieves this task.*

Proof. Note that since Bob has no input, Alice actually knows the message M' in the proof of Theorem 3.1 in this case. Huffman encoding of M' gives the desired protocol, since $H(M') \leq I + \log(I) + O(1)$. □

4 Lower Bound

Now we give an example where Theorem 3.1 is tight. Alice is given a uniformly random string $x \in_R \{0,1\}^n$. Let $M(x,i)$ denote a message distributed according to $x_1, \ldots, x_{i-1}, \bar{x_i}, b_{i+1}, \ldots, b_n$, where b_j's are random bits $\sim B_{1/2}$ and $\bar{x_i}$ denotes the flip of bit x_i.

Given x, Alice's task, T, is to transmit a message distributed according to $M(x,I)$, where $I \in_R \{1,2,\ldots,n\}$. Note that Bob has no input in this task.

First let us bound the private-coin information complexity of this task. Given x, Alice can privately sample I and send $M \sim M(x,I)$. Then the information cost of this protocol is $I(M;X) = H(M) - H(M|X)$. It is clear that $H(M) = n$.

$$H(M|X) = \mathbb{E}_x[H(M|X=x)]$$

Denote $M|X=x$ by $M|_x$. For strings $x,y \in \{0,1\}^n$ with $x \neq y$, let $j(x,y)$ denote the first index of disagreement between x and y i.e. index j s.t. $x_j \neq y_j$. Then

$$Pr[M|_x = y] = \frac{1}{n} \cdot \frac{1}{2^{n-j(x,y)}}$$

if $x \neq y$ and 0 if $x = y$.

$$\begin{aligned} H(M|_x) &= \sum_y Pr[M|_x = y] \log\left(\frac{1}{Pr[M|_x = y]}\right) \\ &= \sum_{j=1}^{n} 2^{n-j} \cdot \frac{1}{n} \cdot \frac{1}{2^{n-j}} \log(n \cdot 2^{n-j}) + 0 \\ &= \log(n) + \frac{1}{n}\sum_{j=1}^{n}(n-j) \\ &= n/2 + \log(n) - 1/2 \end{aligned}$$

The second equality follows from the fact that there are 2^{n-j} strings y with $j(x,y) = j$, when $j \in \{1,\ldots,n\}$. This gives

$$I(M;X) = n/2 - \log(n) + 1/2$$

The following lemma lower bounds the information complexity of a public round protocol for the task T. Note that the strategy of sampling I publicly would have an information cost $\approx n/2$.

Lemma 4.1. *Let Π be a one round public-coin protocol (using public randomness R) such that there is a deterministic function g such that $g(\Pi_x, R)$ is distributed according to $M(x,I)$. Then $I(\Pi; X|R) \geq n/2 - O(1)$.*

Proof. Since Π is a deterministic function of X and R,

$$I(\Pi; X|R) = H(\Pi|R) - H(\Pi|X,R) = H(\Pi|R)$$

Let J be a random variable that denotes the first index of disagreement between $g(\Pi, R)$ and X (Note that J is well defined because of the distribution of M). Fix a value of $R = r$. Let

$p_j = Pr[J = j|R = r]$. Note that the probability is just over random X. Let μ denote the distribution of $\Pi|R = r$ and let μ_j be the distribution of $\Pi|R = r, J = j$. Then

$$\mu = \sum_{j=1}^{n} p_j \cdot \mu_j$$

Let us analyze the distribution μ_j. Let $S_r(j)$ be the set of x's which lead to $J = j$ i.e.

$$S_r(j) = \{x \in \{0,1\}^n : j(x, g(\Pi(x,r), r)) = j\}$$

Note that $|S_r(j)| = p_j \cdot 2^n$. Fixing $\Pi = t$ and $R = r$ fixes $g(\Pi, R) = g(t, r)$. Then

$$Pr[\Pi = t|R = r, J = j] \leq \frac{|\{x \in S_r(j) : j(x, g(t,r) = j)\}|}{|S_r(j)|} \leq \frac{2^{n-j}}{p_j \cdot 2^n} = \frac{1}{p_j \cdot 2^j}$$

The first inequality is because if $R = r, J = j$ are fixed, the event $\Pi = t$ implies that $j(x, g(t,r)) = j$. The second inequality follows from the fact that there are 2^{n-j} x's with $j(x, g(t,r)) = j$.

Claim 4.2. $H(\mu) \geq \sum_{j=1}^{n} j \cdot p_j - O(1)$.

Given the claim, we can bound $H(\Pi|R)$ as follows :

$$\begin{aligned} H(\Pi|R) &= \mathbb{E}_{r\sim R}[H(\Pi|R = r)] \\ &\geq \mathbb{E}_{r\sim R} \sum_{j=1}^{n} j \cdot p_j - O(1) \\ &= \sum_{j=1}^{n} j \cdot \frac{1}{n} - O(1) \\ &= n/2 - O(1) \end{aligned}$$

The inequality follows from the claim. The second equality follows from the fact that $\mathbb{E}_{r\sim R} Pr[J = j|R = r] = Pr[J = j] = \frac{1}{n}$. □

Proof. (Of Claim 4.2) Increasing a larger probability and decreasing a smaller probability by the same amount always lowers the entropy of a distribution

$$\left(p \log\left(\frac{1}{p}\right)\right)' - \left(q \log\left(\frac{1}{q}\right)\right)' = \log\left(\frac{q}{p}\right) < 0 \text{ if } q < p$$

We are given a μ_j where the mass of every entry $\mu_j(z)$ does not exceed $2^{-j}/p_j$. Therefore, we can replace μ_j with a uniform distribution on a set $\mathcal{L}_j$ of L_j entries, where $L_j = \max(1, \lfloor p_j \cdot 2^j \rfloor)$ (given any z_1, z_2 with $0 < \mu_j(z_1), \mu_j(z_2) < 1/L_j$ we can make sure that one of them becomes 0 or that one of them becomes $1/L_j$ without increasing the entropy). Note that it is always the case that $L_j > p_j \cdot 2^{j-1}$.

Therefore, we can assume wlog that each μ_j is uniform on a set $\mathcal{L}_j$ of size L_j. Consider the process of selecting an index K according to the distribution p_j, and then $Z \sim \mu_K$. Our goal is to show that $H(Z) \geq \sum_{j=1}^{n} j \cdot p_j - O(1)$. We have

$$H(KZ) = H(K) + H(Z|K) = \sum_{j=1}^{n} p_j \log(L_j/p_j) > \sum_{j=1}^{n} p_j \log(p_j \cdot 2^{j-1}/p_j) = \sum_{j=1}^{n} j \cdot p_j - 1,$$

and $H(Z) = H(KZ) - H(K|Z)$. Therefore, it suffices to show that $H(K|Z) = O(1)$.
We define a subset S of j's for which p_j is "small":

$$S := \{j : p_j < 2^{-j}\}.$$

Note that for $j \notin S$ we have $p_j \cdot 2^j \geq 1$, and therefore $L_j = \lfloor p_j \cdot 2^j \rfloor$, and $p_j \cdot 2^{j-1} < L_j \leq p_j \cdot 2^j$.
Denote by χ_S the indicator random variable for the event $K \in S$. We have

$$H(K|Z) \leq H(K, \chi_S|Z) = H(\chi_S|Z) + H(K|\chi_S Z) \leq 1 + \Pr[K \in S]H(K|Z, K \in S) + \Pr[K \notin S]H(K|Z, K \notin S).$$

The second inequality is because χ_S is a boolean random variable. We bound the two terms separately. Assuming $S \neq \emptyset$, denote $p_S := \sum_{j \in S} p_j$.

$$\Pr[K \in S]H(K|Z, K \in S) \leq \Pr[K \in S]H(K|K \in S) = p_S \cdot \sum_{j \in S} \frac{p_j}{p_S} \log \frac{p_S}{p_j} \leq \sum_{j \in S} p_j \log \frac{1}{p_j} <$$
$$1 + \sum_{j \geq 2, j \in S} p_j \log \frac{1}{p_j} \leq 1 + \sum_{j=2}^{n} 2^{-j} \log \frac{1}{2^{-j}} = O(1).$$

The last inequality is because the function $x \log 1/x$ is monotone increasing on the interval $(0, 1/e)$, and we have $0 < p_j < 2^{-j} < 1/e$ for $j \in S, j \geq 2$.

Finally, we need to show $\Pr[K \notin S]H(K|Z, K \notin S) = O(1)$. We will in fact show that $H(K|Z, K \notin S) = O(1)$. We have

$$H(K|Z, K \notin S) = \mathbb{E}_{z \sim Z|_{K \notin S}} H(K|Z = z, K \notin S). \tag{1}$$

Fix any value of z such that $\Pr[K \notin S|Z = z] > 0$. We can precisely describe the distribution q of $K|Z = z, K \notin S$. Denote $T_z := \{j : j \notin S, z \in \mathcal{L}_j\}$. Order the elements of T_z in increasing order, and index them: $T_z = \{j_1 < j_2 < \ldots < j_k\}$. Then the distribution q puts weight $q_r := \frac{p_{j_r}/L_{j_r}}{q}$ on j_r, where $q := \sum_{r=1}^{k} p_{j_r}/L_{j_r}$. We have for each r:

$$q_r \leq \frac{p_{j_r}/L_{j_r}}{p_{j_1}/L_{j_1}} < \frac{p_{j_r}/(p_{j_r} \cdot 2^{j_r - 1})}{p_{j_1}/(p_{j_1} \cdot 2^{j_1})} = 2^{j_1 - j_r + 1} \leq 2^{2-r}.$$

The second inequality follows from $L_{j_r} > p_{j_r} \cdot 2^{j_r - 1}$ and $L_{j_1} \leq p_{j_1} \cdot 2^{j_1}$(since $j_1 \notin S$) . $q_r \leq 2^{2-r}$ implies that $H(q) = O(1)$. Therefore we have $H(K|Z = z, K \notin S) = O(1)$ for each z, and by (1) this implies $H(K|Z, K \notin S) = O(1)$, and completes the proof. □

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 67–76, New York, NY, USA, 2010. ACM.

[BBK+12] Joshua Brody, Harry Buhrman, Michal Kouckỳ, Bruno Loff, Florian Speelman, and Nikolay Vereshchagin. Towards a reverse newman's theorem in interactive information complexity. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 19, page 179, 2012.

[BGPW13] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, pages 151–160. ACM, 2013.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *FOCS*, pages 748–757, 2011.

[Bra12] Mark Braverman. Interactive information complexity. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 505–524, New York, NY, USA, 2012. ACM.

[BRWY13] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. *ECCC*, 20(35), 2013.

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):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 Bob Werner, editor, *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, Los Alamitos, CA, October 14–17 2001. IEEE Computer Society.

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of information theory.* Wiley-Interscience, New York, NY, USA, 1991.

[HJMR07] Prahladh Harsha, Rahul Jain, David A. McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. In *IEEE Conference on Computational Complexity*, pages 10–23. IEEE Computer Society, 2007.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, Cambridge, 1997.

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67–71, 31 July 1991.

[Pan12] Denis Pankratov. *Direct sum questions in classical communication complexity.* PhD thesis, Masters thesis, University of Chicago, 2012.

Appendix

Proof. (Of corollary 3.3) It follows by applying Theorem 3.1 to the messages round by round. Denote the protocol transcript by $\Pi = \Pi_1, \Pi_2, \ldots, \Pi_r$. Assume Alice and Bob send alternate messages with Alice sending Π_1. Then

$$\begin{aligned} IC_\mu(\pi) &= I(\Pi;X|YR'R_B)+I(\Pi;Y|XR'R_A) \\ &= \sum_{i\le r} I(\Pi_i;X|Y\Pi_1\Pi_2\dots\Pi_{i-1}R'R_B)+\sum_{i\le r} I(\Pi_i;Y|X\Pi_1\Pi_2\dots\Pi_{i-1}R'R_A) \\ &= \sum_{i \text{ odd}, i\le r} I(\Pi_i;X|Y\Pi_1\Pi_2\dots\Pi_{i-1}R'R_B)+\sum_{i \text{ even}, i\le r} I(\Pi_i;Y|X\Pi_1\Pi_2\dots\Pi_{i-1}R'R_A) \end{aligned}$$

The second equality is chain rule for mutual information and the last equality follows from the fact that for odd i, Π_i is a function of $\Pi_1\Pi_2\dots\Pi_{i-1}$ and X and for even i, Π_i is a function of $\Pi_1\Pi_2\dots\Pi_{i-1}$ and Y. Now, after the messages $\Pi_1=m_1,\Pi_2=m_2,\dots,\Pi_{i-1}=m_{i-1}$ have been sent (assume i odd), Alice can send Π_i using public randomness via a message Π'_i and public randomness R such that (apply Theorem 3.1 to the inputs $XY|\Pi_1=m_1,\Pi_2=m_2\dots\Pi_{i-1}=m_{i-1}$)

$$\begin{aligned} &I(\Pi'_i;X|Y,\Pi_1=m_1,\Pi_2=m_2,\dots,\Pi_{i-1}=m_{i-1},R)\le I(\Pi_i;X|Y\Pi_1=m_1,\Pi_2=m_2\dots\Pi_{i-1} \\ &=m_{i-1}R'R_B)+\log(I(\Pi_i;X|Y\Pi_1=m_1,\Pi_2=m_2,\dots\Pi_{i-1}=m_{i-1}R'R_B))+O(1) \end{aligned}$$

This gives by taking expectations and by concavity of log

$$\begin{aligned} &I(\Pi'_i;X|Y\Pi_1\Pi_2\dots\Pi_{i-1}R)\le I(\Pi_i;X|Y\Pi_1\Pi_2\dots\Pi_{i-1}R'R_B)+\log(I(\Pi_i;X|Y\Pi_1\Pi_2\dots\Pi_{i-1}R'R_B)) \\ &+O(1) \end{aligned}$$

Also by Fact 2.10, $I(\Pi'_i;X|Y\Pi'_1\Pi'_2\dots\Pi'_{i-1}R)=I(\Pi'_i;X|Y\Pi_1\Pi_2\dots\Pi_{i-1}R)$. This is because $\Pi'_1,\dots,\Pi'_{i-1}$, Y, R determine $\Pi_1,\Pi_2,\dots\Pi_{i-1}$ and $\Pi'_1\Pi'_2\dots\Pi'_{i-1}\to YR\Pi_1\Pi_2\dots\Pi_{i-1}\to\Pi'_iX$ is a Markov chain. Thus

$$\begin{aligned} IC_\mu(\pi') \le& \sum_{i\le r, i \text{ odd}} I(\Pi_i;X|Y\Pi_1\Pi_2\dots\Pi_{i-1}R'R_B)+\sum_{i\le r, i\text{ odd}}\log(I(\Pi_i;X|Y\Pi_1\Pi_2\dots\Pi_{i-1}R'R_B))+ \\ &\sum_{i\le r, i\text{ even}} I(\Pi_i;Y|X\Pi_1\Pi_2\dots\Pi_{i-1}R'R_A)+\sum_{i\le r, i\text{ odd}}\log(I(\Pi_i;Y|X\Pi_1\Pi_2\dots\Pi_{i-1}R'R_A))+O(r) \\ &\le I+r\log(I/r)+O(r) \end{aligned}$$

The last inequality follows from concavity of log. □

ECCC ISSN 1433-8092
http://eccc.hpi-web.de