Lower bounds for disjointness using information theory

We follow lecture notes by Mark Braverman (`https://www.cs.princeton.edu/courses/archive/fall11/cos597D/L17.pdf`), fleshing out all the details.

Preliminaries

1 *Disjointness and AND function.*

$$\mathsf{DISJ}(\bar{x}, \bar{y}) = \neg \bigvee_i x_i \wedge y_i$$

2 *KL-divergence and mutual information.*

$$D\left(p \,\|\, q\right) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$I(X : Y) = \mathop{\mathsf{E}}_y [D\left(X|_y \,\|\, X\right)]$$

☞ Let us now use these notions to measure how much *information* (as opposed to *communication*) is revealed by a protocol.

3 *Protocols.* Two-player protocols with public randomness (R) and private randomness (R_a, R_b); worst-case communication; distributional variant.

4 *Information cost.*

$$\mathrm{IC}_\mu(\pi) = I(Y : \Pi | X, R, R_a) + I(X : \Pi | Y, R, R_b)$$

5 *Total variation distance.* (Twice the statistical distance)

$$\|p - q\|_1 = \sum_x |p(x) - q(x)|$$

6 *Pinsker inequality.*

$$\|p-q\|_1 \leq \sqrt{\frac{1}{2} D\left(p \,\|\, q\right)}$$

7 *Pinsker inequality for convex sums.* Suppose that p and q are given by the same convex sum:

$$p(x) = \sum_r \alpha(r) p_r(x) \qquad q(x) = \sum_r \alpha(r) q_r(x),$$

where the $\alpha(r)$ are non-negative reals summing to 1. Then

$$\|p-q\|_1 \leq \sqrt{\frac{1}{2} \sum_r \alpha(r) D\left(p_r \,\|\, q_r\right)}.$$

Proof. From the triangle and Pinkser's inequalities:

$$\|p-q\|_1 \leq \sum_r \alpha_r \|p_r - q_r\|_1 \leq \sum_r \alpha_r \sqrt{\frac{1}{2} D\left(p_r \,\|\, q_r\right)},$$

and then from the concavity of the square-root.

The lower bound

8 **Theorem.** There is no two-player protocol for computing disjointness using $o(n)$ bits of information.

☞ This theorem was originally shown by Kalyanasundaram and Schitger in 1987, and later simplified by Razborov in 1990 via a technique that came to be known as the *corruption bound.* Then in 2004 Bar-Yossef, Jayram, Kumar and Sivakumar prove the theorem via information theory, which allows for a much simpler proof. The proof below was taken from the notes of a lecture by Mark Braverman — I believe it is due to him — and it is even simpler.

9 *Approach.* The proof is split into two parts.

- In the first part, we show that a $o(n)$-information protocol for disjointness would give a $o(1)$-information protocol for the AND function.
- In the second part we show that the latter cannot exist.
- The first part is one of the fundamental techniques of the area. It is essentially a use of the chain rule for mutual information.

- It is obvious that the AND function needs 1 bit of communication in order to be computed by two players when each of their inputs is a uniform independent bit.[1] This is not exactly the statement we need to prove though, as we will see.

First Part

☞ Let $\mu(x,y)$ be the uniform distribution on the support $\{00, 01, 10\}$.

10 Theorem. Let π be a randomized protocol for solving disjointness with success probability $\geq \frac{9}{10}$ using C bits of communication. Then there is a protocol π' for computing the AND function, with the following properties:

10.1 *Correctness.* For all x, y, $\pi'(x; y) = x \wedge y$ with probability at least $\frac{9}{10}$.

10.2 *Low information cost.* The protocol π' has $\leq \frac{2C}{n}$ information cost with respect to the distribution μ, i.e., $\mathrm{IC}_\mu(\pi') \leq \frac{2C}{n}$.

Proof. Define π' as follows. On inputs x and y, Alice and Bob use shared randomness to pick a uniformly-random coordinate $I \in [n]$. They also jointly sample random bits $Y_1, \ldots, Y_{i-1}$ and $X_{i+1}, \ldots, X_n$ to be 0 with probability 2/3 (meaning they pick from the X and Y marginals of μ).

Then Alice privately samples X_j, for $j < i$, so that X_j, Y_j is distributed according to μ, meaning, she lets $X_j = 0$ if $Y_j = 1$, and she lets X_j be a uniformly-random bit if $Y_j = 0$. Bob does the same to privately sample Y_j for $j > i$.

All pairs X_j, Y_j for $j \neq i$ have been defined. The players then set $X_i = x$ and $Y_i = y$, and run the protocol $\pi(\bar{X}; \bar{Y})$.

It now happens (with probability $\geq \frac{9}{10}$) that $\pi(\bar{X}; \bar{Y}) = \mathsf{DISJ}(\bar{X}; \bar{Y}) = \mathsf{NAND}(x, y)$, so after running π both players know $x \wedge y$. Which establishes §10.1.

Now §10.2 follows from the chain rule. Indeed, if X, Y are drawn from μ, then every pair X_j, Y_j is equidistributed. We then have that

$$\frac{C}{n} \geq \frac{1}{n} I(\Pi : \bar{Y} | \bar{X}) = \frac{1}{n} \sum_{j=1}^{n} I(\Pi : Y_j | \bar{X}, Y_{<j}) = I(\Pi : Y_I | \bar{X}, Y_{<I}, I).$$

The rightmost term is exactly the information revealed by π' to Alice about

[1] It follows from the fact that $\mathrm{AND}(x, y)$ reveals $\Omega(1)$ information about x and y, hence by the information processing inequality, the transcript of a protocol for computing AND must also.

Bob's input. Together with the symmetric calculation, this establishes that $\mathrm{IC}_\mu(\pi') \leq \frac{2C}{n}$. ■

Second Part

☞ The lower bound of §8 follows from §10 and the following:

11 Theorem. There is no protocol π for the AND function which is both correct (as in §10.1) and has $\mathrm{IC}_\mu(\pi) = o(1)$.

Proof. First we show that if the information cost of the protocol is $o(1)$, then the transcript distributions for all inputs in the support of μ must be close in total-variation distance.

Indeed, fix some choice for public randomness preserving the information cost. Let R_b denote Bob's private randomness, and suppose that $\alpha(r) = \Pr[R_b = r]$. Then

$$\begin{aligned} o(1) &\geq I(\Pi : X|Y, R_b) \\ &= \frac{2}{3} I(\Pi : X|Y = 0, R_b) + \frac{1}{3} I(\Pi : X|Y = 1, R_b) \\ &= \frac{2}{3} I(\Pi : X|Y = 0, R_b) \\ &= \frac{2}{3} \left(\frac{1}{2} \sum_r \alpha(r) D\left(\Pi_{00r} \,\|\, \Pi_{?0r}\right) + \frac{1}{2} \sum_r \alpha(r) D\left(\Pi_{10r} \,\|\, \Pi_{?0r}\right) \right), \end{aligned}$$

and hence $\sum_r \alpha(r) D\left(\Pi_{00r} \,\|\, \Pi_{?0r}\right)$ and $\sum_r \alpha(r) D\left(\Pi_{10r} \,\|\, \Pi_{?0r}\right)$ are both $o(1)$. From Pinkser's Inequality for convex sums (§7) it then follows that $\|\Pi_{00} - \Pi_{?0}\|_1$ and $\|\Pi_{10} - \Pi_{?0}\|_1$ are both $o(1)$. Now the triangle inequality gives us $\|\Pi_{00} - \Pi_{10}\|_1 = o(1)$.

Doing the same calculation for $I(\Pi : X|Y)$ shows that $\|\Pi_{00} - \Pi_{01}\|_1 = o(1)$, and again by the triangle inequality we also find that $\|\Pi_{10} - \Pi_{01}\|_1 = o(1)$. So the transcript distributions for all inputs in the support of μ is close in total-variation distance.

However, π is also correct on the input $(1, 1)$, which is not on the support of μ, and on which π must output a different result. Taking the error probability into account, we are still forced to conclude that the statistical distance between Π_{00}, say, and Π_{11} is at least $\frac{8}{10}$, meaning $\|\Pi_{00} - \Pi_{11}\|_1 \geq \frac{16}{10}$.

But now we show the following: because π is a protocol, the fact that Π_{00}, Π_{10} and Π_{01} are close to each other must imply that they are also close to Π_{11}. If we denote by $\pi_{xy}(z)$ the probability that $\Pi_{xy} = z$, then it happens that $\pi_{xy}(z) = P_x(z) Q_y(z)$ for some P_x, Q_y. $P_x(z)$ is actually the probability that Alice, when given x, produces a transcript consistent with z. Suppose, for instance, that Alice speaks on the odd-numbered rounds; then

$$P_x(z) = \prod_{i \text{ odd}} \mathsf{Pr}\left[\begin{array}{c}\text{Alice sends bit } z_i \text{ on the } i\text{-th round,}\\ \text{if her input is } x \text{ and she has seen } z_{<i}\end{array}\right].$$

For a given transcript z, suppose that $\pi_{00}(z) \geq \pi_{11}(z)$, and notice the following:

1. If $P_1(z) > P_0(z)$, then $\pi_{11}(z) > \pi_{01}(z)$, and thus $|\pi_{00}(z) - \pi_{11}(z)| < |\pi_{00}(z) - \pi_{01}(z)|$.
2. If $Q_1(z) > Q_0(z)$, then $\pi_{11}(z) > \pi_{10}(z)$, and thus $|\pi_{00}(z) - \pi_{11}(z)| < |\pi_{00}(z) - \pi_{10}(z)|$.
3. If $P_1(z) \leq P_0(z)$ and $Q_1(z) \leq Q_0(z)$, then

   $$(P_0(z) - P_1(z))(Q_0(z) - Q_1(z)) \geq 0,$$

   meaning

   $$-\pi_{11}(z) \leq \pi_{00}(z) - \pi_{10}(z) - \pi_{01}(z),$$

   and then

   $$|\pi_{00}(z) - \pi_{11}(z)| \leq |\pi_{00}(z) - \pi_{10}(z)| + |\pi_{00}(z) - \pi_{01}(z)|.$$

In either case we find that

$$\sum_{z:\pi_{00}(z)\geq\pi_{11}(z)} |\pi_{00}(z) - \pi_{11}(z)| \leq \sum_z |\pi_{00}(z) - \pi_{10}(z)| + \sum_z |\pi_{00}(z) - \pi_{01}(z)|$$

The left-hand side is exactly $\frac{1}{2}$ of $\|\Pi_{00} - \Pi_{11}\|_1$, and the right-hand side is less than $\|\Pi_{00} - \Pi_{10}\|_1 + \|\Pi_{00} - \Pi_{01}\|_1 = o(1)$. Hence $\frac{16}{10} \leq \|\Pi_{00} - \Pi_{11}\|_1 = o(1)$, a contradiction. ■