A Lower Bound for Sampling Disjoint Sets

Mika Göös* Thomas Watson†
Stanford University of Memphis

March 28, 2020

Abstract

Suppose Alice and Bob each start with private randomness and no other input, and they wish to engage in a protocol in which Alice ends up with a set $x \subseteq [n]$ and Bob ends up with a set $y \subseteq [n]$, such that $(x, y)$ is uniformly distributed over all pairs of disjoint sets. We prove that for some constant $\beta < 1$, this requires $\Omega(n)$ communication even to get within statistical distance $1 - \beta^n$ of the target distribution. Previously, Ambainis, Schulman, Ta-Shma, Vazirani, and Wigderson (FOCS 1998) proved that $\Omega(\sqrt{n})$ communication is required to get within some constant statistical distance $\varepsilon > 0$ of the uniform distribution over all pairs of disjoint sets of size $\sqrt{n}$.

1 Introduction

In most traditional computational problems, the goal is to take an input and produce the “correct” output, or produce one of a set of acceptable outputs. In a sampling problem, on the other hand, the goal is to generate a random sample from a specified probability distribution $D$, or at least from a distribution that is close to $D$. There has been a surge of interest in studying sampling problems from a complexity theory perspective [ASTS+03, GGN10, Vio12a, Aar14, LV12, DW12, Vio14, BIL12, Vio12b, JSWZ13, Wat14, BCS14, Wat16, Vio16, Wat18, Vio20]. Unlike more traditional computational problems, sampling problems do not necessarily need to have any real input, besides the uniformly random bits fed into a sampling algorithm.

One commonly studied type of target distribution is “input–output pairs” of a function $f$, i.e., $(D, f(D))$ where $D$ is perhaps the uniform distribution over inputs to $f$. This means an outcome should be $(x, z)$ where $x$ is distributed according to $D$, and $z = f(x)$. Using an algorithm for computing $f$, one can sample $(D, f(D))$ by first sampling from $D$, then evaluating $f$ on that input. However, for some functions $f$, generating an input jointly with the corresponding output may be computationally easier than evaluating $f$ on an adversarially-chosen input. Thus in general, sampling lower bounds tend to be more challenging to prove than lower bounds for functions.

Many of the above-cited works focus on concrete computational models such as low-depth circuits. We consider the model of 2-party communication complexity, for which comparatively less is known about sampling. Which problem should we study? Well, the single most important function in communication complexity is Set-Disjointness, in which Alice gets a set $x \subseteq [n]$, Bob gets a set $y \subseteq [n]$, and they wish to determine whether $x$ and $y$ are disjoint. However, for some functions $f$, generating an input jointly with the corresponding output may be computationally easier than evaluating $f$ on an adversarially-chosen input. Thus in general, sampling lower bounds tend to be more challenging to prove than lower bounds for functions.

*Supported by NSF grant CCF-1412958.
†Supported by NSF grant CCF-1657377.
and at the end Alice outputs some
in the proof was a lemma that was originally employed in [Samp]
are independent), and
this does not matter which party is responsible for outputting the bit
properly-distributed (\(D\)) as a matrix with rows and columns both indexed by \(\{0, 1\}^n\)
where \(D_{x,y}\) is the probability of outcome \((x, y)\). We define \(\text{Samp}(D)\) as the minimum communication cost of any protocol where Alice and Bob each start with private randomness and no other input, and at the end Alice outputs some \(x \in \{0, 1\}^n\) and Bob outputs some \(y \in \{0, 1\}^n\) such that \((x, y)\) is distributed according to \(D\). Note that \(\text{Samp}(D) = 0\) iff \(D\) is a product distribution (\(x\) and \(y\) are independent), and \(\text{Samp}(D) \leq n\) for all \(D\) (since Alice can privately sample \((x, y)\) and send \(y\) to Bob). Allowing public randomness would not make sense since Alice and Bob could read a properly-distributed \((x, y)\) off of the randomness without communicating. We define \(\text{Samp}_2(D)\) as the minimum of \(\text{Samp}(D')\) over all distributions \(D'\) with \(\Delta(D, D') \leq \varepsilon\), where \(\Delta\) denotes statistical (total variation) distance, defined as

\[
\Delta(D, D') := \max_{\text{event } E} \|P_D[E] - P_{D'}[E]\| = \frac{1}{2} \sum_{\text{outcome } o} \|P_D[o] - P_{D'}[o]\|.
\]

1.1 A story

Our story begins with [ASTS+03], which proved that \(\text{Samp}_2((D, \text{Disj}(D))) \geq \Omega(\sqrt{n})\) for some constant \(\varepsilon > 0\), where \(D\) is uniform over the set of all pairs of sets of size \(\sqrt{n}\) (note that this \(D\) is a product distribution and is approximately balanced between 0-inputs and 1-inputs of \(\text{Disj}\)); here it does not matter which party is responsible for outputting the bit \(\text{Disj}(D)\). The main tool in the proof was a lemma that was originally employed in [BFS86] to prove an \(\Omega(\sqrt{n})\) bound on the randomized communication complexity of computing \(\text{Disj}\). The latter bound was improved to \(\Omega(n)\) via several different proofs [KS92, Raz92, BYJKS04], which leads to a natural question: Can we improve the sampling bound of [ASTS+03] to \(\Omega(n)\) by using the techniques of [KS92, Raz92, BYJKS04] instead of [BFS86]?

For starters, the answer is “no” for the particular \(D\) considered in [ASTS+03]—there is a trivial exact protocol with \(O(\sqrt{n} \log n)\) communication since it only takes that many bits to specify a set of size \(\sqrt{n}\). What about other interesting distributions \(D\)? The following illuminates the situation.
Observation 1. For any $D$ and constants $\varepsilon > \delta > 0$, if $\text{Samp}_\varepsilon((D, \text{Disj}(D))) \geq \omega(\sqrt{n})$ then $\text{Samp}_\delta(D) \geq \Omega(\text{Samp}_\varepsilon((D, \text{Disj}(D))))$.

Proof. It suffices to show $\text{Samp}_\varepsilon((D, \text{Disj}(D))) \leq \text{Samp}_\delta(D) + O(\sqrt{n})$. First, note that for any sampling protocol, if we condition on a particular transcript then the output distribution becomes product (Alice and Bob are independent after they stop communicating). Second, [BGK15] proved that for every product distribution and every constant $\gamma > 0$, there exists a deterministic protocol that uses $O(\sqrt{n})$ bits of communication and computes Disj with error probability $\leq \gamma$ on a random input from the distribution. Now to $\varepsilon$-sample $(D, \text{Disj}(D))$, Alice and Bob can $\delta$-sample $D$ to obtain $(x, y)$, and then conditioned on that sampler’s transcript, they can run the average-case protocol from [BGK15] for the corresponding product distribution with error $\varepsilon - \delta$. A simple calculation shows this indeed gives statistical distance $\varepsilon$. \hfill \qed

The upshot is that to get an improved bound, the hardness of sampling $(D, \text{Disj}(D))$ would come entirely from the hardness of just sampling $D$. Thus such a result would not really be “about” the Set-Disjointness function, it would be about the distribution on inputs. Instead of abandoning this line of inquiry, we realize that if $D$ itself is somehow defined in terms of Disj, then a bound for sampling $D$ would still be saying something about the complexity of Set-Disjointness. In fact, the proof in [ASTS+03] actually shows something stronger than the previously-stated result: If $D$ is instead defined as the uniform distribution over pairs of disjoint sets of size $\sqrt{n}$ (which are 1-inputs of Disj), then $\text{Samp}_\varepsilon(D) \geq \Omega(\sqrt{n})$. After this pivot, we are now facing a direction in which we can hope for an improvement. We prove that by removing the restriction on the sizes of the sets, the sampling problem becomes maximally hard. Our result holds for error $\varepsilon < 1$ that is exponentially close to 1, but the result is already new and interesting for constant $\varepsilon > 0$.

Theorem 1. Let $U$ be the uniform distribution over the set of all $(x, y) \in \{0, 1\}^n \times \{0, 1\}^n$ with $x \land y = 0^n$. There exists a constant $\beta < 1$ such that $\text{Samp}_{1-\beta^n}(U) = \Omega(n)$.

The proof from [ASTS+03] was a relatively short application of the technique from [BFS86], but for Theorem 1, harnessing known techniques for proving linear communication lower bounds turns out to be more involved.

For calibration, the uniform distribution over all $(x, y)$ achieves statistical distance $1 - 0.75^n$ from $U$ since there are $4^n$ inputs and $3^n$ disjoint inputs (for a disjoint input, each coordinate $i \in [n]$ has 3 possibilities $x_i, y_i \in \{00, 01, 10\}$). We can do a little better: Suppose for each coordinate independently, Alice picks 0 with probability $\sqrt{1/3}$ and picks 1 with probability $1 - \sqrt{1/3}$, and Bob does the same. This again involves no communication, and it achieves statistical distance $1 - (2\sqrt{1/3} - 1/3)^n \leq 1 - 0.82^n$ from $U$. Theorem 1 shows that the constant 0.82 cannot be improved arbitrarily close to 1 without a lot of communication. (In the setting of lower bounds for circuit samplers, significant effort has gone into handling statistical distances exponentially close to the maximum possible [DW12, BIL12, Vio20].)

1.2 Interpreting the result

As an important step in the proof of Theorem 1, we first observe that our sampling model is equivalent to two other models. One of these we call (for lack of a better word) “synthesizing” the distribution $D$: Alice and Bob get inputs $x, y \in \{0, 1\}^n$ respectively, in addition to their private randomness, and their goal is to accept with probability exactly $D_{x,y}$. We let $\text{Synth}(D)$ denote
the minimum communication cost of any synthesizing protocol for \(D\), and \(\text{Synth}_e(D)\) denote the minimum of \(\text{Synth}(D')\) over all \(D'\) with \(\Delta(D, D') \leq \varepsilon\). The other model is the nonnegative rank of a matrix: \(\text{rank}_+(D)\) is defined as the minimum \(k\) for which \(D\) (viewed as a \(2^n \times 2^n\) matrix) can be written as a sum of \(k\) many nonnegative rank-1 matrices.

**Observation 2.** For every distribution \(D\), the following are all within \(\pm O(1)\) of each other:

\[
\text{Samp}(D), \quad \text{Synth}(D), \quad \log \text{rank}_+(D).
\]

**Proof.** \(\text{Synth}(D) \leq \text{Samp}(D) + 2\) since a synthesizing protocol can just run a sampling protocol and accept iff the result equals the given input \((x, y)\). (Only this part of Observation 2 is needed in the proof of Theorem 1.)

\[
\log \text{rank}_+(D) \leq \text{Synth}(D)
\]

since for each transcript of a synthesizing protocol, the matrix that records the probability of getting that transcript on each particular input has rank 1 (since Alice’s private randomness being consistent with the transcript, and Bob’s private randomness being consistent with the transcript, are independent events); summing these matrices over all accepting transcripts yields a nonnegative rank decomposition of \(D\).

To see that \(\text{Samp}(D) \leq \lceil \log \text{rank}_+(D) \rceil\), suppose \(D = M^{(1)} + M^{(2)} + \cdots + M^{(k)}\) is a sum of nonnegative rank-1 matrices. For each \(i\), by scaling we can write \(M^{(i)}_{x,y} = p_i u^{(i)}_x v^{(i)}_y\), for some distributions \(u^{(i)}\) and \(v^{(i)}\) over \([0,1]^n\), where \(p_i\) is the sum of all entries of \(M^{(i)}\). Since \(D\) is a distribution, \(p := (p_1, \ldots, p_k)\) is a distribution over \([k]\). To sample from \(D\), Alice can privately sample \(i \sim p\) and send it to Bob using \(\lceil \log k \rceil\) bits, then Alice can sample \(x \sim u^{(i)}\) and Bob can independently sample \(y \sim v^{(i)}\) with no further communication. \(\square\)

By this characterization, Theorem 1 can be viewed as a lower bound on the approximate nonnegative rank of the DISJ matrix, where the approximation is in \(\ell_1\) (which has an average-case flavor). In the recent literature, “approximate nonnegative rank” generally refers to approximation in \(\ell_\infty\) (which is a worst-case requirement), and this model is equivalent to the so-called smooth rectangle bound and WAPP communication complexity [JK10, KMSY19, GLM+16].

Observation 2 combined with a result of [LS93] shows that the deterministic communication complexity of any total two-party boolean function \(f\) is quadratically related to the communication complexity of exactly sampling the uniform distribution over \(f^{-1}(1)\).

# 2 Proof

## 2.1 Overview

Our proof of Theorem 1 is by a black-box reduction to the well-known corruption lemma for Set-Disjointness due to Razborov [Raz92]. We start with a high-level overview.

For notation: Let \(|z|\) denote the Hamming weight of a string \(z \in \{0,1\}^n\). For \(\ell \in \mathbb{N}\), let \(U^\ell\) be the uniform distribution over all \((x, y) \in \{0,1\}^n \times \{0,1\}^n\) with \(|x \wedge y| = \ell\). Note that \(U = U^0\). For a distribution \(D\) over \([0,1]^n \times [0,1]^n\) and an event \(E \subseteq [0,1]^n \times [0,1]^n\), let \(D_E := \sum_{(x,y) \in E} D_{x,y}\).

For a randomized protocol \(\Pi\), let \(\text{acc}_\Pi(x, y)\) denote the probability that \(\Pi\) accepts \((x, y)\).

**Step I: Uniform corruption.** The corruption lemma states that if a rectangle \(R \subseteq \{0,1\}^n \times \{0,1\}^n\) contains a noticeable fraction of disjoint pairs, then it must contain about as large a fraction
of uniquely intersecting pairs. More quantitatively, there exist a constant $C > 0$ and two distributions $D^1, \ell = 0, 1$, defined over disjoint $(\ell = 0)$ and uniquely intersecting pairs $(\ell = 1)$ such that for every rectangle $R$,

$$\text{if } D^0_R \geq 2^{-o(n)} \text{ then } D^1_R \geq C \cdot D^0_R.$$  

The original proof [Raz92] defined $D^\ell$ as the uniform distribution over all pairs $(x, y)$ with fixed sizes $|x| = |y| = \lceil n/4 \rceil$ and $|x \land y| = \ell$. For our purpose, we need the corruption lemma to hold relative to the aforementioned distributions $U^\ell, \ell = 0, 1$, which have no restrictions on set sizes. We derive in §2.2 a corruption lemma for $U^\ell$ from the original lemma for $D^\ell$. To do this, we exhibit a reduction that uses public randomness and no communication to transform a sample from $D^\ell$ into a sample from a distribution that is close to $U^\ell$ in a suitable sense, for $\ell = 0, 1$.

**Step II: Truncate and scale.** For simplicity, let us think about proving Theorem 1 for a small error $\varepsilon > 0$. Assume for contradiction there is some distribution $D$, $\Delta(U, D) \leq \varepsilon$, such that $\text{Synth}(D) \leq o(n)$ as witnessed by a private-randomness synthesizing protocol $\Pi'$ with $\text{acc}_{\Pi'}(x, y) = D_{x,y}$. Note that the total acceptance probability over disjoint inputs is close to 1:

$$\sum_{x,y : |x \land y| = 0} \text{acc}_{\Pi'}(x, y) \geq 1 - \varepsilon \text{ and thus } \mathbb{E}_{(x,y) \sim U^0}[\text{acc}_{\Pi'}(x, y)] \geq (1 - \varepsilon)3^{-n}.$$  

Our eventual goal (in Step III) is to apply our corruption lemma to the transcript rectangles, but the above threshold $(1 - \varepsilon)3^{-n}$ is too low for this. To raise the threshold to $2^{-o(n)}$ as needed for corruption, we would like to scale up all the acceptance probabilities accordingly. To “make room” for the scaling, we first carry out a certain truncation step. Specifically, in §2.3 we transform $\Pi'$ into a public-randomness protocol $\Pi$:

1. First, we **truncate** (using a truncation lemma [GLM+16]) the values $\text{acc}_{\Pi'}(x, y)$, which has the effect of decreasing some of them, but any $\text{acc}_{\Pi'}(x, y)$ that is under $3^{-n}$ remains approximately the same. This results in an intermediate protocol $\Pi''$ that still satisfies $\mathbb{E}_{(x,y) \sim U^0}[\text{acc}_{\Pi''}(x, y)] \geq \Omega((1 - \varepsilon)3^{-n})$ (using the assumption that $\Delta(U, D) \leq \varepsilon)$.

2. Second, we **scale** (using the low cost of $\Pi''$) the truncated probabilities up by a large factor $3^n2^{-o(n)}$. This results in a protocol $\Pi$ with large typical acceptance probabilities:

$$\mathbb{E}_{(x,y) \sim U^0}[\text{acc}_{\Pi}(x, y)] \geq 2^{-o(n)}. \quad (1)$$  

**Step III: Iterate corruption.** Because $\Pi$ has such large acceptance probabilities (1), our corruption lemma can be applied: there is some constant $C' > 0$ such that

$$\mathbb{E}_{(x,y) \sim U^1}[\text{acc}_{\Pi}(x, y)] \geq C' \cdot \mathbb{E}_{(x,y) \sim U^0}[\text{acc}_{\Pi}(x, y)]. \quad (2)$$  

Since $\Pi$ is a truncated-and-scaled version of $\Pi'$, this allows us to infer that

$$\mathbb{E}_{(x,y) \sim U^1}[\text{acc}_{\Pi'}(x, y)] \geq \Omega((1 - \varepsilon)3^{-n}) \text{ and thus } \sum_{x,y : |x \land y| = 1} \text{acc}_{\Pi'}(x, y) \geq \Omega((1 - \varepsilon)n)$$  

using the fact that $|\text{supp}(U^1)| = n3^{n-1} = (n/3) \cdot |\text{supp}(U^0)|$. Thus for $\varepsilon = 1 - \omega(1/n)$, this means $\Pi'$ must have placed a total probability mass $> 1$ on uniquely intersecting inputs, which is the sought contradiction.

To prove Theorem 1 for very large error $\varepsilon = 1 - \beta^n$, in §2.4 we iterate the above argument for $U^\ell$ over $0 \leq \ell \leq o(n)$. Namely, analogously to (2), we show that the average acceptance
probability of \( \Pi \) over \( U^{\ell+1} \) is at least a constant times the average over \( U^\ell \). Meanwhile, the support sizes increase as \( |\text{supp}(U^{\ell+1})| \geq \omega(1) \cdot |\text{supp}(U^\ell)| \) for \( \ell = o(n) \). These facts together imply a large constant factor increase in the total probability mass that \( \Pi' \) places on \( \text{supp}(U^{\ell+1}) \) as compared to \( \text{supp}(U^\ell) \). Starting with even a tiny probability mass over \( \text{supp}(U^0) \), this iteration will eventually lead to a contradiction.

### 2.2 Step I: Uniform Corruption

The goal of this step is to derive Lemma 2 from Lemma 1.

**Lemma 1 (Corruption [Raz92]).** For every rectangle \( R \subseteq \{0,1\}^n \times \{0,1\}^n \) we have \( D_R^1 \geq \frac{1}{75} D_R^0 - 2^{-0.017n} \) where, assuming \( n = 4k - 1 \), \( D_0^\ell \) is the uniform distribution over all \((x,y)\) with \(|x| = |y| = k\) and \(|x \wedge y| = \ell\).

**Lemma 2 (Uniform Corruption).** For every rectangle \( R \subseteq \{0,1\}^n \times \{0,1\}^n \) we have \( U_R^1 \geq \frac{1}{75} U_R^0 - 2^{-0.008n} \).

**Proof.** Assume for convenience that \( n/2 \) has the form \( 4k - 1 \) (otherwise use the nearest such number instead of \( n/2 \) throughout). We prove that Lemma 1 for \( n/2 \) implies Lemma 2 for \( n \) by the contrapositive. Thus, \( D_0^R \) and \( D_1^R \) are distributions over \( \{0,1\}^{n/2} \times \{0,1\}^{n/2} \) while \( U_0^R \) and \( U_1^R \) are distributions over \( \{0,1\}^n \times \{0,1\}^n \). Assume there exists a rectangle \( R \subseteq \{0,1\}^n \times \{0,1\}^n \) such that \( U_R^1 < \frac{1}{75} U_R^0 - 2^{-0.008n} \). We exhibit a distribution over rectangles \( Q \subseteq \{0,1\}^{n/2} \times \{0,1\}^{n/2} \) such that \( \mathbb{E}[D_Q^1] < \frac{1}{75} \mathbb{E}[D_Q^0] - 2^{-0.017n/2} \); by linearity of expectation this implies that there exists such a \( Q \) with \( D_Q^1 < \frac{1}{75} D_Q^0 - 2^{-0.017n/2} \).

To this end, we define a distribution \( F \) over functions \( f: \{0,1\}^{n/2} \times \{0,1\}^{n/2} \rightarrow \{0,1\}^n \times \{0,1\}^n \) of the form \( f(x,y) = (f_1(x), f_2(y)) \) and then let \( Q_f \) be the rectangle \( f^{-1}(R) := \{(x,y) : f(x,y) \in R\} \). Let \( H \) be the distribution over \( \{(v,w) \in \mathbb{N} \times \mathbb{N} : v + w \leq n\} \) obtained by sampling \( (x,y) \sim U_0^R \) and outputting \((|x|, |y|)\); i.e., \( H_{v,w} := \frac{n!}{v!w!(n-v-w)!} \cdot 3^{-n} \). To sample \( f \sim F \):

1. Sample \((v,w)\) from \( H \) conditioned on \( v \geq k, w \geq k, \) and \( v + w \leq 2k + n/2 \).
2. Sample a uniformly random permutation \( \pi \) of \([n]\).
3. Given \((x,y) \in \{0,1\}^{n/2} \times \{0,1\}^{n/2}\), define \((x', y') \in \{0,1\}^n \times \{0,1\}^n \) by letting

\[
\begin{align*}
x'_i y'_i &:= \\
&\begin{cases}
x_i y_i & \text{for the first } n/2 \text{ coordinates } i; \\
10 & \text{for the next } v - k \text{ coordinates } i; \\
01 & \text{for the next } w - k \text{ coordinates } i; \\
00 & \text{for the remaining } n/2 - (v - k) - (w - k) \geq 0 \text{ coordinates } i.
\end{cases}
\end{align*}
\]

4. Let \( f(x,y) := (\pi(x'), \pi(y')) \) (i.e., permute the coordinates according to \( \pi \)).

For \( \ell \in \{0,1\} \) let \( F(D_\ell) \) denote the distribution obtained by sampling \((x,y) \sim D_\ell \) and \( f \sim F \) and outputting \( f(x,y) \), and note that \( F(D_\ell)_R = \mathbb{E}_F[D_\ell^{R_{Q_F}}] \). Now we claim that \( F(D_\ell) \) and \( U_\ell^R \) are close, in the following senses:

1. For every event \( E \), \( F(D^0)_E \geq U^0_E - 2^{-0.01n} \).
2. For every event \( E \), \( F(D^1)_E \leq U^1_E \cdot 17 \).
Using $R$ as the event $E$, we have

$$F(D^1)_R \leq U_R^1 \cdot 17$$

$$< 17 \left( \frac{1}{100} U_R^0 - 2^{-0.008n} \right)$$

$$\leq 17 \left( \frac{1}{100} \left(F(D^0)_R + 2^{-0.01n} \right) - 2^{-0.008n} \right)$$

$$\leq 1.5 F(D^0)_R - 2^{-0.017n/2}$$

as desired. To see (1), note that $F(D^0)$ is precisely $U^0$ conditioned on $v \geq k$, $w \geq k$, and $v + w \leq 2k + n/2$, and this conditioning event has probability at least $1 - 2^{-0.01n}$ by Chernoff bounds:

$$\mathbb{P}[v < k] = \mathbb{P}[w < k] = \mathbb{P}[\text{Bin}(n, 1/3) < n/8 + 1/4] \leq 2^{-0.12n}$$

$$\mathbb{P}[v + w > 2k + n/2] = \mathbb{P}[\text{Bin}(n, 2/3) > 3n/4 + 1/2] \leq 2^{-0.02n}.$$ 

Thus letting $C$ be the complement of the conditioning event, we have $F(D^0)_E \geq U^0_E \cdot C \geq U^0_E - U^0_C \geq U^0_E - 2^{-0.01n}$. To see (2), consider any outcome $(x, y) \in \{0, 1\}^n \times \{0, 1\}^n$ with $|x \land y| = 1$. We have $U^1_{x,y} = 1/(n3^{n-1})$. Abbreviating $a := |x|$ and $b := |y|$, assume $a \geq k$, $b \geq k$, and $a + b \leq 2k + n/2$ since otherwise $F(D^1)_{x,y} = 0$ and there would be nothing to prove. Henceforth consider the probability space with the randomness of $D^1$ and of $F$. Let $I$ be the event that $F_1(D^1) \land F_2(D^1) = x \land y$, i.e., that the intersecting coordinate of $F(D^1)$ is the same as for $(x, y)$. We have

$$F(D^1)_{x,y} = \mathbb{P}[I] \cdot \mathbb{P}[v = a \text{ and } w = b] \cdot \mathbb{P}[F(D^1) = (x, y) \mid I \text{ and } v = a \text{ and } w = b].$$

For the three terms on the right side, we have

$$(*) = \frac{1}{n}, \quad (**) \leq H_{a,b}/(1-2^{-0.01n}) \leq \frac{n!}{a!b!(n-a-b)!} \cdot 3^{-n} \cdot 1.01, \quad (***) = 1/\left( (a-1)!(b-1)!(n-a-b+1)! \right).$$

We have

$$\frac{n!}{a!b!(n-a-b)!} / \left( (a-1)!(b-1)!(n-a-b+1)! \right) = \frac{n-(n-a-b+1)}{a \cdot b} \leq \frac{n-(n-2k+1)}{k \cdot k} \leq \frac{n-(n-2n/8+1)}{(n/8) \cdot (n/8)} = \left( \frac{3}{4} + \frac{1}{n} \right) \cdot 64.$$ 

Combining, we get

$$F(D^1)_{x,y} / U^1_{x,y} = (\ast) \cdot (\ast\ast) \cdot (\ast\ast\ast) \cdot n3^{n-1} \leq \frac{1.01}{3} \cdot \left( \frac{3}{4} + \frac{1}{n} \right) \cdot 64 \leq 17. \quad \square$$

### 2.3 Step II: Truncate and scale

The goal of this step is to construct a truncated-and-scaled protocol $\Pi$ from any given low-cost $\Pi'$ that synthesizes a distribution close to $U$.

For a nonnegative matrix $M$, we define its truncation $\overline{M}$ to be the same matrix but where each entry $> 1$ is replaced with 1.

**Lemma 3 (Truncation Lemma [GLM+16]).** For every $2^n \times 2^n$ nonnegative rank-1 matrix $M$ and every natural number $d$, there exists a $O(d + \log n)$-communication public-randomness protocol $\Pi$ such that for every $(x, y)$ we have $\text{acc}_\Pi(x, y) \in \overline{M}_{x,y} \pm 2^{-d}$. 

7
Let $c \geq 1$ be the hidden constant in the big $O$ in Lemma 3, and let $\delta := 0.00005/c$. Toward proving Theorem 1, suppose for contradiction $\text{Samp}(D) \leq \delta n$ for some distribution $D$ with $\Delta(U, D) \leq 1 - 2^{-\delta n}$ (so $\beta := 2^{-\delta}$ in Theorem 1) and thus

$$\sum_{x,y : |x\cap y| = 0} \min(3^{-n}, D_{x,y}) = \sum_{x,y} \min(U_{x,y}, D_{x,y})$$

$$= \sum_{x,y} U_{x,y} - \sum_{x,y : U_{x,y} > D_{x,y}} (U_{x,y} - D_{x,y})$$

$$= 1 - \Delta(U, D)$$

$$\geq 2^{-\delta n}.$$ 

By Observation 2, $\text{Synth}(D) \leq \delta n + 2$, so consider a synthesizing protocol $\Pi'$ for $D$ with communication cost $\leq \delta n + 2$. Let $A$ be the set of all accepting transcripts of $\Pi'$. For each $\tau \in A$ let $N^\tau$ be the nonnegative rank-1 matrix such that $N^\tau_{x,y}$ is the probability $\Pi'$ generates $\tau$ on input $(x, y)$; thus $D_{x,y} = \sum_{\tau \in A} N^\tau_{x,y}$. Let $\Pi^\tau$ be the public-randomness protocol from Lemma 3 applied to $M^\tau := 3^n N^\tau$ and $d := 15\delta n$. Let $\Pi$ be the public-randomness protocol that picks a uniformly random $\tau \in A$ and then runs $\Pi^\tau$. The communication cost of $\Pi$ is $\leq c \cdot (d + \log n) \leq 0.001n$.

**Claim 1.** For every input $(x, y)$ we have $\frac{3^n}{|A|} \min(3^{-n}, D_{x,y}) - 2^{-d} \leq \text{acc}_\Pi(x, y) \leq \frac{3^n}{|A|} D_{x,y} + 2^{-d}$.

**Proof.** We have

$$\text{acc}_\Pi(x, y) = \frac{1}{|A|} \sum_{\tau \in A} \text{acc}_{\Pi^\tau}(x, y)$$

$$\leq \frac{1}{|A|} \sum_{\tau \in A} (\tau_{x,y} + 2^{-d})$$

$$\leq \frac{1}{|A|} \sum_{\tau \in A} \min(1, 3^n N^\tau_{x,y}) + 2^{-d}$$

$$= \frac{3^n}{|A|} \sum_{\tau \in A} \min(3^{-n}, N^\tau_{x,y}) + 2^{-d}.$$ 

From this it follows that:

$$\text{acc}_\Pi(x, y) \geq \frac{3^n}{|A|} \min(3^{-n}, \sum_{\tau \in A} N^\tau_{x,y}) - 2^{-d} = \frac{3^n}{|A|} \min(3^{-n}, D_{x,y}) - 2^{-d}$$

$$\text{acc}_\Pi(x, y) \leq \frac{3^n}{|A|} \sum_{\tau \in A} N^\tau_{x,y} + 2^{-d} = \frac{3^n}{|A|} D_{x,y} + 2^{-d}. \hspace{1cm} \square$$

We can now formally state the large typical acceptance probability property (equation (1) from the overview): writing $U_\Pi := \mathbb{E}(x,y) \sim U[\text{acc}_\Pi(x, y)]$ (and similarly for other input distributions),

$$U_\Pi \geq \frac{1}{|A|} \sum_{x,y : |x\cap y| = 0} \left( \frac{3^n}{|A|} \min(3^{-n}, D_{x,y}) - 2^{-d} \right) \hspace{1cm} \text{(by Claim 1)}$$

$$\geq \frac{1}{|A|} 2^{-\delta n} - 2^{-15\delta n}$$

$$\geq \frac{1}{|A|} 2^{-\delta n - 1} \hspace{1cm} \text{(3)}$$

where the last line follows because $|A| \leq 2^{\delta n + 2}$ and $2^{-2\delta n - 2}$ is at least twice $2^{-15\delta n}$.

**2.4 Step III: Iterate corruption**

Here we derive the final contradiction: $\Pi'$ places an acceptance probability mass exceeding 1 on $\text{supp}(U^{\delta n})$. This is achieved by iterating our corruption lemma, starting with (3) as the base case.
For \( z \in \{0,1\}^n \) let \( U^z \) be the uniform distribution over all \((x,y) \in \{0,1\}^n \times \{0,1\}^n\) with \( x \land y = z \) (so \( U^\ell \) is the uniform mixture of all \( U^z \) with \(|z| = \ell\); in particular, \( U^0 = U^{0n} \)), and if \(|z| < n\) then let \( \widehat{U}^z \) be the uniform mixture of \( U^{z'} \) over all \( z' \) that can be obtained from \( z \) by flipping a single 0 to 1 (so \( U^{\ell+1} \) is the uniform mixture of all \( \widehat{U}^z \) with \(|z| = \ell\); in particular, \( U^1 = \widehat{U}^{0n} \)).

**Claim 2.** For every \( z \in \{0,1\}^n \) with \(|z| \leq n/2\) we have \( \widehat{U}^z_{\Pi} \geq \frac{1}{\sqrt{65}} U^z_{\Pi} - 2^{-0.003n} \).

**Proof.** Since all relevant inputs \((x,y)\) have \( x_i y_i = 11\) for all \( i \) such that \( z_i = 1\), we can ignore those coordinates and think of \( \widehat{U}^z \) and \( U^z \) as \( U^1 \) and \( U^0 \) respectively, but defined on the remaining \( n-|z| \geq n/2 \) coordinates (instead of on all \( n \) coordinates). Thus by Lemma 2, for every outcome of the public randomness of \( \Pi \) and every accepting transcript, say corresponding to rectangle \( R \), we have \( \widehat{U}^z_R \geq \frac{1}{\sqrt{65}} U^z_R - 2^{-0.008n/2} \). Summing over all the (at most \( 2^{0.001n} \) many) accepting transcripts, and then taking the expectation over the public randomness, yields the claim since \( 2^{0.001n} \cdot 2^{-0.008n/2} \leq 2^{-0.003n} \).

**Claim 3.** For every \( \ell = 0,\ldots,\delta n \) we have \( U^\ell_{\Pi} \geq \frac{1}{|A|} 2^{-\delta n - 1 - 11\ell} \).

**Proof.** We prove this by induction on \( \ell \). The base case \( \ell = 0 \) is (3). For the inductive step, assume the claim is true for \( \ell \). Since \( U^{\ell+1} \) and \( U^\ell \) are the uniform mixtures of \( \widehat{U}^z \) and \( U^z \) respectively over all \( z \) with \(|z| = \ell \) (so \( U^{\ell+1} = E_z[\widehat{U}^z] \) and \( U^\ell = E_z[U^z] \)), by linearity of expectation Claim 2 implies

\[
U^\ell_{\Pi} \geq \frac{1}{\sqrt{65}} U^\ell_{\Pi} - 2^{-0.003n} \geq \frac{1}{|A|} 2^{-\delta n - 1 - 11\ell - \log_2(765)} \geq 2^{-0.003n} \geq \frac{1}{|A|} 2^{-\delta n - 1 - 11(\ell+1)}
\]

where the last inequality follows because \(|A| \leq 2^{\delta n+2} \) and \( 2^{-\delta n - 2 - \delta n - 1 - 11\delta n - \log_2(765)} \geq 2^{-14\delta n} \) is at least twice \( 2^{-0.003n} \), which gives \( U^{\ell+1}_{\Pi} \geq \frac{1}{|A|} 2^{-\delta n - 1 - 11\ell - \log_2(765) - 1} \), and \( \log_2(765) + 1 \leq 11 \).

Choosing \( \ell = \delta n \) we have

\[
U^\ell_{\Pi} \geq 2^{-\delta n - 2 - \delta n - 1 - 11\delta n} \geq 2^{-15\delta n} \]

because \(|A| \leq 2^{\delta n+2} \) and \( 2^{-\delta n - 2 - \delta n - 1 - 11\delta n} \geq 2^{-14\delta n} \) is at least twice \( 2^{-15\delta n} \). Thus, for \( \ell = \delta n \),

\[
\sum_{x,y} D_{x,y} \geq \sum_{x,y: |x \land y| = \ell} D_{x,y} \geq \sum_{x,y: |x \land y| = \ell} \frac{|A|}{n^\ell} (\text{acc}_1(x,y) - 2^{-d})
\]

(by Claim 1)

\[
= \frac{|A|}{n^\ell} \binom{n}{\ell} 3^{n-\ell} (U^\ell_{\Pi} - 2^{-d})
\]

\[
\geq \frac{|A|}{n^\ell} \binom{n}{\ell} 3^{n-\ell} \frac{1}{|A|} 2^{-\delta n - 2 - 11\ell}
\]

\[
= \left( \frac{n^{\ell + 2}}{3^{-11\ell}} \right) 2^{-\delta n - 2 - \ell / 2}
\]

\[
= \left( \frac{\delta + 2}{\delta - 3} \right)^{\delta n / 2}
\]

\[
\geq 1.6^{\delta n}
\]

\[
> 1,
\]

contradicting the fact that \( D \) is a distribution.
A Information complexity proof

In this appendix we provide an alternative proof of a weaker version of Theorem 1 that only handles statistical distance 0.01 instead of $1 - e^{-\beta n}$. This proof may be of independent interest, and it is somewhat more self-contained than the proof of Theorem 1 since it does not rely on corruption.

**Theorem 2 (Weaker version of Theorem 1).** Let $U$ be the uniform distribution over the set of all $(x, y) \in \{0, 1\}^n \times \{0, 1\}^n$ with $x \land y = 0^n$. Then $\text{Samp}_{0.01}(U) = \Omega(n)$.

A.1 Overview

We use the Synth characterization from Observation 2 in our proof of Theorem 2. We also use the information complexity method that was pioneered in [CSWY01, BYJKS04] for proving the randomized $\Omega(n)$ bound for computing Disj. At a high level, the information complexity approach is to consider a probability space with a random input (from a distribution of our choosing) and with the random transcript generated by a protocol on that input, and to use the fact that the communication cost is lower bounded by the Shannon entropy of the transcript, which in turn is lower bounded by the “information cost”: the mutual information between the transcript and the input. The key is that as long as the $n$ input coordinates are independent of each other, the information cost obeys a direct sum property: It is at least the sum of the contributions of the $n$ coordinates. Thus an $\Omega(n)$ bound follows by showing that the mutual information between the transcript and a typical coordinate is $\Omega(1)$.

How shall we implement this approach, given a synthesizing protocol for a distribution that is close to the $U$ from Theorem 2? First we should decide which input distribution to measure information cost with respect to. For this we use $U$ itself, but the reason is not only because the aim is to prove a lower bound for approximately synthesizing $U$. We use $U$ also because statistical (or $\ell_1$) distance is a certain sum—rather than a weighted sum—over inputs.

When looking at an individual coordinate’s contribution to the information cost, we need the input to come from a product distribution, in order to exploit the fact that each transcript corresponds to a combinatorial rectangle. The standard technique is to decompose the input distribution into a mixture of product distributions (like what a sampling protocol does—but now this is purely for analysis purposes) and consider a typical component of this mixture. Then, we can use a standard lemma (from [BYJKS04]) for relating the mutual information to the statistical distance between transcript distributions on different inputs; however, we need to somewhat generalize this tool to handle the input distributions that arise from decomposing our $U$.

The next issue to tackle is that a synthesizing protocol rejects most inputs with extremely high probability, so the rejecting transcripts may not carry much information about the input (the information cost could be very low if we measure w.r.t. a random transcript in the naive way). Most of the “action” happens within the approximately $3^{-n}$ probability of acceptance on a typical 1-input. For this reason, our probability space will use a random transcript conditioned on the protocol accepting. This introduces two related sources of difficulties: It distorts the “product structure” we usually rely on for analyzing the behavior of a transcript across different inputs, and it interferes with the standard trick of “absorbing” the other $n - 1$ coordinates of the randomly chosen input into the protocol’s private randomness (specifically, sampling an input and then a random accepting transcript on that input, is not the same as sampling an input, running the protocol, then conditioning the whole experiment on acceptance). A substantial portion of the technical effort goes into alleviating these issues.
Anyway, to give the gist of the overall structure of the argument, let us visualize a single “representative” transcript and ignore the complications mentioned in the previous paragraph. Focusing on a single input coordinate (Disj with \( n = 1 \) is just \textsc{Nand}), we think of the lower-right cell as a 0-input and the other three cells as 1-inputs. The area within each cell represents the protocol’s private randomness along with “the rest” of the random input (besides the coordinate under the spotlight). If the transcript’s rectangle occupies too much area in the upper-left cell (as shown on the left), it would be contributing to the protocol accepting 1-inputs with too high of probability. Otherwise, the rectangle is forced to occupy a relatively not-too-small area in the lower-right cell (as shown on the right). Accepting on some 0-inputs can be OK, but here is the catch: There are \( n/3 \) times as many uniquely-intersecting inputs as there are disjoint inputs (for the full \textsc{Disj} function), and it turns out this not-too-small acceptance probability would get “replicated” across many of these intersecting inputs. The sum of the acceptance probabilities would then be too great for the protocol to be synthesizing any distribution at all, much less one close to \( U \).

![Diagram of Disj function](image)

A.2 Preliminaries

We assume familiarity with the basics of communication complexity [KN97] and information theory [CT06]. A protocol \( \Pi \) is assumed to have private randomness, and we let \( \text{CC}(\Pi) \) denote the worst-case communication cost. We use \( \mathbb{P} \) for probability, \( \mathbb{E} \) for expectation, \( \mathbb{H} \) for Shannon entropy, \( \mathbb{I} \) for mutual information, \( \mathbb{D} \) for relative entropy, and \( \Delta \) for statistical distance. We use bold letters to denote random variables, and non-bold letters for particular outcomes.

**Fact 1.** Mutual information and relative entropy satisfy the following standard properties:

- **Direct sum:** \( \mathbb{I}(\mathbf{a} ; \mathbf{b}_1 \cdots \mathbf{b}_n) \geq \mathbb{I}(\mathbf{a} ; \mathbf{b}_1) + \cdots + \mathbb{I}(\mathbf{a} ; \mathbf{b}_n) \) if \( \mathbf{b}_1 \cdots \mathbf{b}_n \) are fully independent.

- **Alternative definition:** \( \mathbb{I}(\mathbf{a} ; \mathbf{b}) = \mathbb{E}_{\mathbf{b} \sim \mathbf{b}} \mathbb{D}(\mathbb{P}(\mathbf{a} | \mathbf{b} = \mathbf{b}) \parallel \mathbb{P}(\mathbf{a})) \).

- **Pinsker’s inequality:** \( \mathbb{D}(\mathbf{a} \parallel \mathbf{b}) \geq \frac{1}{2m^2} \Delta(\mathbf{a}, \mathbf{b})^2 \).

Here is the quick proof of the direct sum property: \( \mathbb{H}(\mathbf{b}_1 \cdots \mathbf{b}_n) = \mathbb{H}(\mathbf{b}_1) + \cdots + \mathbb{H}(\mathbf{b}_n) \) by full independence, and \( \mathbb{H}(\mathbf{b}_1 \cdots \mathbf{b}_n | \mathbf{a}) \leq \mathbb{H}(\mathbf{b}_1 | \mathbf{a}) + \cdots + \mathbb{H}(\mathbf{b}_n | \mathbf{a}) \) by subadditivity of entropy. Thus

\[
\mathbb{I}(\mathbf{a} ; \mathbf{b}_1 \cdots \mathbf{b}_n) = \mathbb{H}(\mathbf{b}_1 \cdots \mathbf{b}_n) - \mathbb{H}(\mathbf{b}_1 \cdots \mathbf{b}_n | \mathbf{a}) \geq \sum_i (\mathbb{H}(\mathbf{b}_i) - \mathbb{H}(\mathbf{b}_i | \mathbf{a})) = \sum_i \mathbb{I}(\mathbf{a} ; \mathbf{b}_i).
\]

Pinsker’s inequality has several proofs available in several sources, such as [DP09].

We also need the following tool relating statistical distance and mutual information. The special case where \( \mathbf{b} \) is uniform over \( \{0, 1\} \) was known, dating back to [BYJKS04] (using [Lin91]). For the general case, we provide a simple proof that was suggested by an anonymous reviewer.
Lemma 4. Let \(a, b\) be jointly distributed, with \(b\) having support \(\{0, 1\}\). Then
\[
\Delta((a \mid b = 0), (a \mid b = 1)) \leq \sqrt{\mathbb{I}(a ; b) / \min(\mathbb{P}[b = 0], \mathbb{P}[b = 1])}.
\]

Proof. By the alternative definition in Fact 1, we have
\[
\mathbb{I}(a ; b) = \mathbb{E}_{b \sim b} \mathbb{D}((a \mid b = b) \parallel a) \geq \min(\mathbb{P}[b = 0], \mathbb{P}[b = 1]) \cdot \sum_{b \in \{0, 1\}} \mathbb{D}((a \mid b = b) \parallel a).
\]

By Pinsker’s inequality in Fact 1 and Cauchy–Schwarz, we have
\[
\sum_{b \in \{0, 1\}} \mathbb{D}((a \mid b = b) \parallel a) \geq \frac{2}{\ln 2} \sum_{b \in \{0, 1\}} \Delta((a \mid b = b), a)^2 \\
\geq \frac{2}{\ln 2} (\sum_{b \in \{0, 1\}} \Delta((a \mid b = b), a))^2 / 2 \\
= \frac{1}{\ln 2} \Delta((a \mid b = 0), (a \mid b = 1))^2.
\]

Combining and using \(\ln 2 \leq 1\) yields the lemma.

\[\square\]

A.3 Proof of Theorem 2

Letting \(U\) be as in Theorem 2, suppose for contradiction \(\text{Samp}(D) \leq 0.0001n - 2\) for some distribution \(D\) with \(\Delta(U, D) \leq 0.01\). By Observation 2, \(\text{Synth}(D) \leq 0.0001n\), so consider a synthesizing protocol \(\Pi\) for \(D\) with \(\text{CC}(\Pi) \leq 0.0001n\). As a technical convenience, we may assume \(\Pi\) has been infinitesimally perturbed to ensure the acceptance probability is positive on each input;\(^1\) this allows us to avoid special cases for handling conditioning on 0-probability events.

We build a probability space by decomposing \(U\) into a mixture of product distributions. For each \(j \in [n]\) independently: Let \(w_j\) be uniform over \(\{\text{LEFT}, \text{RIGHT}\}\).

Conditioned on \(w_j = \text{LEFT}\), let \(x_jy_j := \begin{cases} 
00 & \text{with probability } 1/3 \\
10 & \text{with probability } 2/3
\end{cases}\).

Conditioned on \(w_j = \text{RIGHT}\), let \(x_jy_j := \begin{cases} 
00 & \text{with probability } 1/3 \\
01 & \text{with probability } 2/3
\end{cases}\).

Note that the marginal distribution of \((x, y)\) is \(U\), but \(x\) and \(y\) are independent conditioned on \(w\). Conditioned on \((x, y) = (x_j, y_j)\), let \(\tau\) be a random transcript of \(\Pi(x, y)\) conditioned on acceptance. Finally, let \(i\) be uniform over \([n]\) and independent of the other random variables. In summary, \((w, x, y, \tau, i)\) are jointly distributed over \(\{\text{LEFT, RIGHT}\}^n \times \{0, 1\}^n \times \{0, 1\}^n \times \{0, 1\}^{\text{CC}(\Pi)} \times [n]\).

Definition 1. An outcome \((i, w_{-i}) \in [n] \times \{\text{LEFT, RIGHT}\}^{[n] \setminus \{i\}}\) is called good iff both:

1. \(\mathbb{I}(\tau ; x_iy_i \mid i = i, w = w) \leq 0.0008\) ("cost")
   for each \(w_i \in \{\text{LEFT, RIGHT}\}\), and
2. \(\mathbb{E}[(3^{-n} - D_{x_iy_i, 0}) \mid i = i, w_{-i} = w_{-i}, x_iy_i = x_iy_i] \leq 0.12 \cdot 3^{-n}\) ("correctness")
   for each \(x_iy_i \in \{00, 01, 10\}\).

Claim 4. \((i, w_{-i}) \sim (i, w_{-i})\) is good with probability at least 0.5.

\(^1\)For an infinitesimal \(\epsilon > 0\), accept with probability \(4^{-\epsilon}\), reject with probability \((1 - 4^{-\epsilon})\), and otherwise run the original synthesizing protocol. This adds only 2 bits of communication, and it affects the statistical distance to the target distribution by at most \(\epsilon\).
Proof. By a union bound, it suffices to show that (1) and (2) individually hold with probability at least 0.75 each.

For fixed \( i \) and \( w \), abbreviate \( \mathbb{I}(\tau; x_i y_i \mid i = i, w = w) \) as \( I_{i,w} \) for \( I_{i,w} \geq 0 \). For each \( w \),

\[
\mathbb{E}_{i \sim I_{i,w}}[I_{i,w}] \leq \frac{1}{n} \mathbb{I}(\tau; x y \mid w = w) \leq \frac{1}{n} \mathbb{H}(\tau \mid w = w) \leq \frac{1}{n} \text{CC}(\Pi) \leq 0.0001
\]

by the first bullet from Fact 1 using \( a := (\tau \mid w = w) \) and \( b_i := (x_i y_i \mid w = w) \). Now

\[
\mathbb{E}_{i \sim w} \mathbb{E}_{i \sim I_{i,w}}[I_{i,w}] = \mathbb{E}_{w \sim w} \mathbb{E}_{i \sim I_{i,w}}[I_{i,w}] \leq 0.0001.
\]

By Markov’s inequality, with probability at least 0.75 over \((i, w_{-i}) \sim (i, w_{-i})\) we have that \( \mathbb{E}_{w_{-i} \sim w_{-i}}[I_{i,w}] \leq 0.0004 \), in which case \( \max_{w_{-i}}(I_{i,w}) \leq 0.0008 \) and thus (1) holds.

For fixed \( i, w_{-i} \), and \( x_i y_i \), abbreviate \( \mathbb{E}[\max(3^n - D_{x,y}, 0) \mid i = i, w_{-i} = w_{-i}, x_i y_i = x_i y_i] \) as \( \delta_{i,w_{-i},x_i y_i} \geq 0 \). Now

\[
\mathbb{E}_{(i, w_{-i}) \sim (i, w_{-i})} \mathbb{E}_{x_i y_i \sim x_i y_i}[\delta_{i,w_{-i},x_i y_i}] = \mathbb{E}[\max(3^n - D_{x,y}, 0) \mid i = i, w_{-i} = w_{-i}, x_i y_i = x_i y_i] = 3^n \Delta(U, D) \leq 0.01 \cdot 3^n.
\]

By Markov’s inequality, with probability at least 0.75 over \((i, w_{-i}) \sim (i, w_{-i})\) we have that \( \mathbb{E}_{x_i y_i \sim x_i y_i}[\delta_{i,w_{-i},x_i y_i}] \leq 0.04 \cdot 3^n \), in which case \( \max_{x_i y_i}[\delta_{i,w_{-i},x_i y_i}] \leq 3 \mathbb{E}_{x_i y_i \sim x_i y_i}[\delta_{i,w_{-i},x_i y_i}] \leq 0.12 \cdot 3^n \) and thus (2) holds.

\[\square\]

Lemma 5. For each good \((i, w_{-i})\), either:

(i) \( \mathbb{E}[D_{x,y} \mid i = i, w_{-i} = w_{-i}] \geq 5 \cdot 3^{-n} \), or

(ii) \( \mathbb{E}[D_{\hat{x},\hat{y}} \mid i = i, w_{-i} = w_{-i}] \geq 0.000001 \cdot 3^{-n} \)

where the random variables \( \hat{x}, \hat{y} \) are the same as \( x, y \) except \( \hat{x}_i \hat{y}_i \) is fixed to 11.

Lemma 5 is the technical heart of the argument; we prove it in §A.4. Note that the marginal distribution of \((\hat{x}, \hat{y})\) is uniform over the set of all \((x, y) \in \{0, 1\}^n \times \{0, 1\}^n\) with \(|x \land y| = 1\), where \(| \cdot |\) denotes Hamming weight (i.e., \( x \) and \( y \) represent uniquely intersecting sets).

Combining Claim 4 and Lemma 5 shows that over \((i, w_{-i}) \sim (i, w_{-i})\), either (i) holds with probability at least 0.25 or (ii) holds with probability at least 0.25. In the former case,

\[
\mathbb{E}[D_{x,y}] \geq \mathbb{P}(\text{(i) holds}) \cdot \mathbb{E}[D_{x,y} \mid \text{(i) holds}] \geq 0.25 \cdot 5 \cdot 3^{-n} > 3^{-n}
\]

and thus \( \sum_{x,y:|x \land y| = 0} D_{x,y} = 3^n \mathbb{E}[D_{x,y}] > 1 \). In the latter case,

\[
\mathbb{E}[D_{\hat{x},\hat{y}}] \geq \mathbb{P}(\text{(ii) holds}) \cdot \mathbb{E}[D_{\hat{x},\hat{y}} \mid \text{(ii) holds}] \geq 0.25 \cdot 0.000001 \cdot 3^{-n} > 1/(n3^{n-1})
\]

and thus \( \sum_{x,y:|x \land y| = 1} D_{x,y} = n3^{n-1} \mathbb{E}[D_{\hat{x},\hat{y}}] > 1 \). Either case contradicts the assumption that \( D \) is a distribution. This finishes the proof of Theorem 2, except for the proof of Lemma 5.

A.4 Proof of Lemma 5

Fix a good \((i, w_{-i})\). For convenience, we henceforth assume \( i = 1 \) and we elide the conditioning on \( i = 1, w_{-1} = w_{-1} \) in the notation. Thus our whole probability space now consists of \((w_1, x, y, \tau)\) which is actually distributed as \((w_1, x, y, \tau \mid i = 1, w_{-1} = w_{-1})\) in the original notation. Also, \((\hat{x}, \hat{y}) := (1x_{-1}, 1y_{-1})\).

With this convention, the definition of good becomes
(1) \( I(\tau ; x_1 y_1 \mid w_1 = w_1) \leq 0.0008 \) \( \text{("cost")} \)
for each \( w_1 \in \{\text{LEFT, RIGHT}\} \), and
\( \mathbf{E} \left[ \max(3^{-n} - D_{x,y}, 0) \mid x_1 y_1 = x_1 y_1 \right] \leq 0.12 \cdot 3^{-n} \) \( \text{("correctness")} \)
for each \( x_1 y_1 \in \{00, 01, 10\} \)

and the statement of \textbf{Lemma 5} becomes

(i) \( \mathbf{E}[D_{x,y}] \geq 5 \cdot 3^{-n} \), or
(ii) \( \mathbf{E}[D_{x,y}] \geq 0.000001 \cdot 3^{-n} \).

\textbf{Claim 5.} \textit{If (1) holds then} \( \Delta((\tau \mid x_1 y_1 = 00), (\tau \mid x_1 y_1 = x_1 y_1)) \leq 0.05 \)
for each \( x_1 y_1 \in \{01, 10\} \).

\textbf{Proof.} First, \( (1) \) tells us \( I(\tau ; x_1 \mid w_1 = \text{LEFT}) \leq 0.0008 \) (since \( y_1 \) is always 0 conditioned on \( w_1 = \text{LEFT} \)), and applying \textbf{Lemma 4} with \((a, b) := (\tau, x_1 \mid w_1 = \text{LEFT})\) gives
\[
\Delta((\tau \mid x_1 y_1 = 00), (\tau \mid x_1 y_1 = 10)) \leq \sqrt{I(\tau ; x_1 \mid w_1 = \text{LEFT})/3} \leq \sqrt{0.0008 \cdot 3} \leq 0.05.
\]
Similarly, \( (1) \) tells us \( I(\tau ; y_1 \mid w_1 = \text{RIGHT}) \leq 0.0008 \) (since \( x_1 \) is always 0 conditioned on \( w_1 = \text{RIGHT} \)), and applying \textbf{Lemma 4} with \((a, b) := (\tau, y_1 \mid w_1 = \text{RIGHT})\) gives
\[
\Delta((\tau \mid x_1 y_1 = 00), (\tau \mid x_1 y_1 = 01)) \leq \sqrt{I(\tau ; y_1 \mid w_1 = \text{RIGHT})/3} \leq \sqrt{0.0008 \cdot 3} \leq 0.05.
\]
This proves the claim. \hfill \( \square \)

Let \( A \) be the set of all accepting transcripts of \( \Pi \). For \( \tau \in A \) and \((x, y) \in \{0, 1\}^n \times \{0, 1\}^n \) and \( x_1 y_1 \in \{0, 1\}^2 \), define
\[
p_{x,y}(\tau) := \mathbb{P}[\Pi(x, y) \text{ generates } \tau] \\
q_{x,y}(\tau) := \mathbb{P}[\Pi(x, y) \text{ generates } \tau \mid \Pi(x, y) \text{ accepts}]
\]
where the probabilities in the left column are over the private randomness of Alice and Bob; in particular, \( \sum_{\tau \in A} p_{x,y}(\tau) = \mathbb{P}[\Pi(x, y) \text{ accepts}] = D_{x,y} \) and \( p_{x,y}(\tau) = D_{x,y} \cdot q_{x,y}(\tau) \). In terms of our probability space \((w_1, x, y, \tau)\), we have:

for each \( x_1 y_1 \in \{00, 01, 10\} \):
\[
\begin{align*}
p_{x_1 y_1}(\tau) &= \mathbb{E} [p_{x,y}(\tau) \mid x_1 y_1 = x_1 y_1] \\
p_{11}(\tau) &= \mathbb{E} [p_{x,y}(\tau)] \\
p_{01}(\tau) &= \mathbb{E} [p_{x,y}(\tau) \mid x_1 y_1 = x_1 y_1] \\
p_{10}(\tau) &= \mathbb{E} [p_{x,y}(\tau)] \\
p_{00}(\tau) &= \mathbb{E} [p_{x,y}(\tau) \mid x_1 y_1 = x_1 y_1]
\end{align*}
\]

where \( \tau \) is always 0 conditioned on \( w_1 = \text{LEFT} \), and \( \tau \) is always 0 conditioned on \( w_1 = \text{RIGHT} \).

We postpone the proofs of the following two claims to the end of this subsection.

\textbf{Claim 6.} \textit{If (2) holds then} \( \mathbb{P}[p_{x_1 y_1}(\tau)/q_{x_1 y_1}(\tau) \geq 0.03 \cdot 3^{-n}] \mid x_1 y_1 = x_1 y_1 \geq 0.8 \)
for each \( x_1 y_1 \in \{01, 10\} \).

\textbf{Claim 7.} \textit{If (i) does not hold then} \( \mathbb{P}[p_{00}(\tau)/q_{00}(\tau) \leq 75 \cdot 3^{-n}] \mid x_1 y_1 = 00] \geq 0.8 \).
We now show how to combine Claim 5, Claim 6, and Claim 7 to prove that if (1) and (2) hold and (i) does not hold, then (ii) holds. Defining

\[ T_{x_1y_1} := \{ \tau \in A : p_{x_1y_1}(\tau)/q_{x_1y_1}(\tau) \geq 0.03 \cdot 3^{-n} \} \quad \text{for each } x_1y_1 \in \{01,1\} \]
\[ T_{00} := \{ \tau \in A : p_{00}(\tau)/q_{00}(\tau) \leq 75 \cdot 3^{-n} \} \quad \text{for } x_1y_1 = 00 \]
\[ T := T_{00} \cap T_{01} \cap T_{10} \]

we have for each \( x_1y_1 \in \{01,1\} \),

\[ \mathbb{P}[\tau \in T_{x_1y_1} | x_1y_1 = 00] \geq \mathbb{P}[\tau \in T_{x_1y_1} | x_1y_1 = x_1y_1] - 0.05 \geq 0.75 \]

by Claim 5 and Claim 6, and \( \mathbb{P}[\tau \in T_{00} | x_1y_1 = 00] \geq 0.8 \) by Claim 7, so by a union bound,

\[ \sum_{\tau \in T} q_{00}(\tau) = \mathbb{P}[\tau \in T | x_1y_1 = 00] \geq 0.3. \tag{†} \]

For each \( x_1y_1 \in \{01,1\} \) we define \( d_{x_1y_1}(\tau) := |q_{00}(\tau) - q_{x_1y_1}(\tau)| \) so that by Claim 5,

\[ \sum_{\tau \in A} d_{x_1y_1}(\tau) = 2 \Delta((\tau | x_1y_1 = 00), (\tau | x_1y_1 = x_1y_1)) \leq 0.1. \tag{‡} \]

Since \( x_{-1}, y_{-1} \) are independent (implicitly conditioned on \( w_{-1} = w_{-1} \)), we have \( p_{00}(\tau) \cdot p_{11}(\tau) = p_{01}(\tau) \cdot p_{10}(\tau) \) by the rectangular nature of any transcript \( \tau \). We would like to rewrite this as \( p_{11}(\tau) = p_{01}(\tau) \cdot p_{10}(\tau) / p_{00}(\tau) \) but we must be careful about division by 0. Adopting the convention \( 0/0 := 0 \), we can write

\[ p_{11}(\tau) \geq p_{01}(\tau) \cdot p_{10}(\tau) / p_{00}(\tau). \tag{*} \]

We also note that for each \( x_1y_1, p_{x_1y_1}(\tau) = 0 \) iff \( q_{x_1y_1}(\tau) = 0 \). To convert between the “multiplicative” structure of transcripts as in \((*)\) and the “additive” structure of statistical distance, we appeal to the following basic fact, which has been used several times in recent works [GW16, GPW16, GJW18]. For completeness, we reproduce the argument at the end of this subsection.

**Fact 2.** For every \( \tau \in A \), \( q_{01}(\tau) \cdot q_{10}(\tau) / q_{00}(\tau) \geq q_{00}(\tau) - d_{01}(\tau) - d_{10}(\tau) \).

At last we come to the punchline:

\[ \mathbb{E}[D_{\tilde{x}, \tilde{y}}] = \sum_{\tau \in A} p_{11}(\tau) \geq \sum_{\tau \in T} p_{11}(\tau) \geq \sum_{\tau \in T} p_{01}(\tau) \cdot p_{10}(\tau) / p_{00}(\tau) \]
\[ = \sum_{\tau \in T} \frac{p_{01}(\tau) \cdot p_{10}(\tau) / p_{00}(\tau)}{q_{00}(\tau) / q_{00}(\tau)} \cdot \frac{q_{01}(\tau) \cdot q_{10}(\tau)}{q_{00}(\tau)} \]
\[ \geq \sum_{\tau \in T} \frac{(0.03 \cdot 3^{-n}) \cdot (0.03 \cdot 3^{-n})}{75 \cdot 3^{-n}} \cdot (q_{00}(\tau) - d_{01}(\tau) - d_{10}(\tau)) \]
\[ \geq 0.00001 \cdot 3^{-n} \cdot \left( \sum_{\tau \in T} q_{00}(\tau) - \sum_{\tau \in A} d_{01}(\tau) - \sum_{\tau \in A} d_{10}(\tau) \right) \]
\[ \geq 0.00001 \cdot 3^{-n} \cdot (0.3 - 0.1 - 0.1) = 0.000001 \cdot 3^{-n} \]

where the third line uses Fact 2, and the last line follows by (†) and (‡). Thus (ii) holds. This finishes the proof of Lemma 5, except for the proofs of Claim 6, Claim 7, and Lemma 4.
Proof of Claim 6. To slightly declutter notation, we write the argument for \(x_1y_1 = 01\) (nothing is different for \(x_1y_1 = 10\)). Assuming (2) holds, we have \(\mathbb{E}[\max(3^{-n} - D_{x,y}, 0) \mid x_1y_1 = 01] \leq 0.12 \cdot 3^{-n}\). We define \(S\) as the set of all \((x, y)\) in the support of \((x, y)\) conditioned on \(x_1y_1 = 01\) (and implicitly on \(w_{-1} = w_{-1}\)) such that \(D_{x,y} \leq 0.16 \cdot 3^{-n}\) (“bad inputs”). By Markov’s inequality,
\[
\mathbb{P}[(x, y) \in S \mid x_1y_1 = 01] = \mathbb{P}[\max(3^{-n} - D_{x,y}, 0) \geq 0.84 \cdot 3^{-n} \mid x_1y_1 = 01] \leq 1/7 \leq 0.15.
\]
We define \(B\) as the set of all \(\tau \in A\) such that \(\mathbb{P}[(x, y) \in S \mid \tau = \tau, x_1y_1 = 01] \geq 0.8\) (“bad transcripts”). We must have \(\mathbb{P}[\tau \in B \mid x_1y_1 = 01] \leq 0.2\) since otherwise
\[
\mathbb{P}[(x, y) \in S \mid x_1y_1 = 01] \geq \mathbb{P}[(x, y) \in S \text{ and } \tau \in B \mid x_1y_1 = 01] = \mathbb{P}[\tau \in B \mid x_1y_1 = 01] \cdot \mathbb{P}[(x, y) \in S \mid \tau \in B, x_1y_1 = 01] \geq 0.2 \cdot 0.8 = 0.16 > 0.15.
\]
Let \(\chi_{x,y}\) be the indicator for \((x, y) \notin S\), so \(D_{x,y} \geq 0.16 \cdot 3^{-n} \cdot \chi_{x,y}\). For each \(\tau \in A \setminus B\) we have
\[
\mathbb{E}[(\chi_{x,y} \cdot q_{x,y}(\tau)) \mid x_1y_1 = 01] = \sum_{(x,y) \notin S} \mathbb{P}[xy = xy \mid x_1y_1 = 01] \cdot \mathbb{P}[\tau = \tau \mid xy = xy] = \mathbb{P}[xy \notin S \text{ and } \tau = \tau \mid x_1y_1 = 01] = \mathbb{P}[(x, y) \notin S \mid \tau = \tau, x_1y_1 = 01] \cdot \mathbb{P}[\tau = \tau \mid x_1y_1 = 01] \geq 0.2 \cdot q_{01}(\tau)
\]
and thus
\[
p_{01}(\tau) = \mathbb{E}[p_{x,y}(\tau) \mid x_1y_1 = 01] = \mathbb{E}[D_{x,y} \cdot q_{x,y}(\tau) \mid x_1y_1 = 01] \geq 0.16 \cdot 3^{-n} \cdot \mathbb{E}[\chi_{x,y} \cdot q_{x,y}(\tau) \mid x_1y_1 = 01] \geq 0.16 \cdot 3^{-n} \cdot 0.2 \cdot q_{01}(\tau) \geq 0.03 \cdot 3^{-n} \cdot q_{01}(\tau).
\]
In summary, \(\mathbb{P}[p_{01}(\tau)/q_{00}(\tau) \geq 0.03 \cdot 3^{-n} \mid x_1y_1 = 01] \geq \mathbb{P}[\tau \notin B \mid x_1y_1 = 01] \geq 0.8\).

Proof of Claim 7. Assume \(\mathbb{E}[D_{x,y} \mid x_1y_1 = 00] \leq 15 \cdot 3^{-n}\) since otherwise (i) would hold because
\[
\mathbb{E}[D_{x,y}] = \mathbb{E}_{x_1y_1 \sim x_1y_1} \mathbb{E}[D_{x,y} \mid x_1y_1 = x_1y_1] \geq \frac{1}{3} \mathbb{E}[D_{x,y} \mid x_1y_1 = 00] \geq 5 \cdot 3^{-n}.
\]
Now
\[
\mathbb{E}\left[\frac{p_{00}(\tau)}{q_{00}(\tau)} \mid x_1y_1 = 00\right] = \sum_{\tau \in A} q_{00}(\tau) \cdot \frac{p_{00}(\tau)}{q_{00}(\tau)} = \sum_{\tau \in A} p_{00}(\tau) = \mathbb{E}[D_{x,y} \mid x_1y_1 = 00] \leq 15 \cdot 3^{-n}.
\]
Thus \(\mathbb{P}[p_{00}(\tau)/q_{00}(\tau) \leq 75 \cdot 3^{-n} \mid x_1y_1 = 00] \geq 0.8\) follows by Markov’s inequality.

Proof of Fact 2. It suffices to show that
\[
q_{01}(\tau) \cdot q_{10}(\tau) \geq q_{00}(\tau)^2 - q_{00}(\tau)(d_{01}(\tau) + d_{10}(\tau)). \tag{5}
\]
(If \(q_{00}(\tau) \neq 0\) then the desired inequality follows by dividing (5) by \(q_{00}(\tau)\), and if \(q_{00}(\tau) = 0\) then it follows since its right side is \(\leq 0\) and its left side is \(0\); recall our convention that \(0/0 := 0\).) For some signs \(\sigma_{x_1y_1}(\tau) \in \{1, -1\}\), the left side of (5) equals \((q_{00}(\tau) + \sigma_{01}(\tau)d_{01}(\tau)) \cdot (q_{00}(\tau) + \sigma_{10}(\tau)d_{10}(\tau))\), which expands to
\[
q_{00}(\tau)^2 + \sigma_{01}(\tau)q_{00}(\tau)d_{01}(\tau) + \sigma_{10}(\tau)q_{00}(\tau)d_{10}(\tau) + \sigma_{01}(\tau)\sigma_{10}(\tau)d_{01}(\tau)d_{10}(\tau).
\]
If \(\sigma_{01}(\tau) = \sigma_{10}(\tau)\) then (6) is at least the right side of (5) since the last term of (6) is nonnegative. If \(\sigma_{01}(\tau) \neq \sigma_{10}(\tau)\), say \(\sigma_{01}(\tau) = -1\) and \(\sigma_{10}(\tau) = 1\), then (6) is at least the right side of (5) since the sum of the last two terms in (6) is \(q_{00}(\tau)d_{10}(\tau) - d_{01}(\tau)d_{10}(\tau) = q_{01}(\tau)d_{10}(\tau) \geq 0\).
Acknowledgments

We thank anonymous reviewers for helpful comments. A preliminary version of this paper was published as [GW19].

References


