THEORY OF COMPUTING, Volume 11 (1), 2015, pp. 1–34
www.theoryofcomputing.org

The Complexity of Deciding Statistical Properties of Samplable Distributions

Thomas Watson*

Received December 10, 2013; Revised February 13, 2015; Published March 12, 2015

Abstract: We consider the problems of deciding whether the joint distribution sampled by a given circuit has certain statistical properties such as being i. i. d., being exchangeable, being pairwise independent, having two coordinates with identical marginals, having two uncorrelated coordinates, and many other variants. We give a proof that simultaneously shows all these problems are $\mathsf{C_=P}$-complete, by showing that the following promise problem (which is a restriction of all the above problems) is $\mathsf{C_=P}$-complete: Given a circuit, distinguish the case where the output distribution is uniform and the case where every pair of coordinates is neither uncorrelated nor identically distributed. This completeness result holds even for samplers that are depth-3 circuits.

We also consider circuits that are d-local, in the sense that each output bit depends on at most d input bits. We give linear-time algorithms for deciding whether a 2-local sampler's joint distribution is fully independent, and whether it is exchangeable.

We also show that for general circuits, certain approximation versions of the problems of deciding full independence and exchangeability are SZK-complete.

We also introduce a bounded-error version of $\mathsf{C_=P}$, which we call $\mathsf{BC_=P}$, and we investigate its structural properties.

ACM Classification: F.1.3

AMS Classification: 68Q17, 68Q15

Key words and phrases: complexity theory, algorithms, complexity classes, completeness, samplable distributions

---

An extended abstract of this paper appeared in the Proceedings of the 31st International Symposium on Theoretical Aspects of Computer Science (STACS), 2014 [39].

*Supported by funding from NSERC.

 DOI: 10.4086/toc.2015.v011a001

1 Introduction

Testing for independence of random variables is a fundamental problem in statistics. Theoretical computer scientists have studied this and other analogous problems from two main viewpoints. The first viewpoint is property testing of distributions, which is a black-box model in which a tester is given samples and tries to distinguish between some statistical property being "close" or "far" from satisfied. Some important works giving upper and lower bounds for property testing of distributions include [6, 5, 4, 7, 29, 2, 28, 35, 30, 26, 14, 13].

The other viewpoint is the white-box model in which a tester is given a description of a distribution (from which it could generate its own samples). This could potentially make some problems easier, but there are complexity-theoretic results showing that several such problems are computationally hard, particularly when the input is a *succinct* description of a distribution. One of the most general and natural ways to succinctly specify a distribution is to give the code of an efficient algorithm that takes "pure" randomness and transforms it into a sample from the distribution. (This gives a polynomial-size specification of a distribution over a potentially exponential-size set.) For arbitrary circuit samplers, the papers [31, 20, 21, 40] contain completeness results for various approximation problems concerning statistical distance, Shannon entropy, and min-entropy. See [22] for a survey of both the black-box and the white-box viewpoints.

In this paper we consider a wide array of "exact" problems concerning statistical properties of the joint distribution produced by a given sampler. Such problems include deciding whether the joint distribution is i. i. d., exchangeable, pairwise independent, and many other variants. Exchangeability is a very important and useful concept with many different applications in pure and applied probability [25, 1], but it has been less-often studied in the theoretical computer science community. A joint distribution over a finite domain is called *exchangeable* if it is invariant under permuting the coordinates. It is fairly straightforward to see that a finite distribution is exchangeable iff it is a mixture of distributions that arise from drawing a sequence of colored balls without replacement from an urn[1] [16]. When each coordinate is a single bit, exchangeability is equivalent to the probability of a string only depending on the Hamming weight. We feel it is natural to pose complexity-theoretic questions about exchangeability.

We prove that the aforementioned wide array of problems, and more generally a single problem we call PANOPTIC-STATS which is at most as hard as any of those problems, are complete for the complexity class $\mathsf{C_=P}$. This class was introduced in [38] as part of the counting hierarchy, and it can be viewed as a class that captures "exact counting" of NP witnesses. The class $\mathsf{C_=P}$ is at least as hard as the polynomial-time hierarchy, since $\mathsf{PH} \subseteq \mathsf{BP} \cdot \mathsf{C_=P}$ [33] and even $\mathsf{PH} \subseteq \mathsf{ZP} \cdot \mathsf{C_=P}$ [32]. It is at most as hard as "threshold counting," since $\mathsf{C_=P} \subseteq \mathsf{PP}$, and it is not substantially easier, since $\mathsf{PP} \subseteq \mathsf{NP}^{\mathsf{C_=P}}$. The class $\mathsf{C_=P}$ has been given several names and characterizations; it equals the classes[2] coNQP [18] and ES [11].

In many areas of complexity theory, when arbitrary small-size circuits are too unwieldy to reason about, we restrict our attention to more stringent complexity measures that are combinatorially simple enough to reason about and obtain unconditional results. The two major categories of such complexity

[1] This is a finite analogue of De Finetti's Theorem.

[2] NQP is a nondeterministic version of quantum polynomial time. It differs from QMA, an alternative such version, in that NQP is defined in terms of zero vs. non-zero probability of acceptance by a quantum algorithm (like the nondeterminism-based definition of NP), whereas QMA is defined in terms of a quantum algorithm given a quantum state as a witness (like the verification-based definition of NP). These two classes are not known to be comparable.

measures are parallel time, and space. One model of efficient parallel time computation is AC^0 (constant-depth unbounded fan-in circuits with AND, OR, and NOT gates). Papers that study AC^0 circuits that sample distributions include [36, 27, 37, 8]. Another (generally more restrictive) model of efficient parallel time computation is *locally-computable* functions, where each output bit depends on at most a bounded number of input bits. Papers that study locally-computable functions as samplers include [36, 17, 15, 37, 40] as well as a large collection of papers investigating the possibility of implementing pseudorandom generators locally. (See [15] for an extensive list of past work on the power of locally-computable functions, including whether they can implement PRGs, one-way functions, and extractors.) The most common model for logarithmic-space samplers is one with streaming/one-way access to the pure random input bits. Topics that have been studied concerning such logspace samplers include compression [34], extraction [24], and min-entropy estimation [40]. One more paper worth mentioning is [10], which considers Markov random fields as succinct descriptions of distributions (though these descriptions would not be considered "samplers").

We prove that our $\mathsf{C_=P}$-completeness results hold even when restricted to samplers that are AC^0-type circuits with depth 3 and top fan-in 2 (i. e., each output gate has fan-in at most 2). We also consider 2-local samplers (where each output bit depends on at most 2 of the pure random input bits) such that each coordinate of the sampled joint distribution is a single bit. We give polynomial-time (in fact, linear-time) algorithms for deciding whether such a sampler's distribution is fully independent, and whether it is exchangeable. These seem to be the first-of-a-kind algorithmic results on deciding statistical properties of succinctly described distributions.

We also consider approximate versions of the problems discussed above: deciding whether the joint distribution of a given sampler is statistically close to or far from satisfying a property. It was shown in [20] that for the property of being uniform, the problem is complete for the class NISZK (non-interactive statistical zero-knowledge). It was shown in [31] that the problem of deciding whether a pair of samplable distributions are statistically close or far is complete for the class SZK (statistical zero knowledge). We prove that with suitable parameters, the approximate versions of the full independence and exchangeability problems (for general circuit samplers) are also SZK-complete.

In this paper we also consider a "bounded-error" version of $\mathsf{C_=P}$, which we call $\mathsf{BC_=P}$ and which does not seem to have been defined or studied in the literature before. Although it does not appear to be directly relevant to statistical properties of samplable distributions, we take the opportunity to study this class and prove that it is closed under several operations (disjunction, conjunction, union, and intersection).

2 Results

If D is a joint distribution over $(\{0,1\}^k)^n$, we let D_i (for $i \in \{1,\dots,n\}$) denote the i^{th} coordinate, which is marginally distributed over $\{0,1\}^k$. For each of the computational problems we consider, the input is a circuit $S : \{0,1\}^r \to (\{0,1\}^k)^n$ (and we assume that the values of k and n are part of the description of the circuit). We call such a circuit a (k,n)-sampler, and if it has size $\le s$ we also call it a (k,n,s)-sampler. Plugging a uniformly random string into S yields a joint output distribution, which we denote by $S(U)$. We formulate computational problems using the framework of promise problems. Throughout this paper, when we talk about reductions and completeness, we are always referring to Karp reductions (polynomial-time many-one reductions). We refer to the texts [3, 19] for expositions of standard complexity classes and completeness.

We state our completeness results for exact problems in Section 2.1 and prove them in Section 3. We state our algorithmic results for exact problems in Section 2.2 and prove them in Section 4. We state our completeness results for approximate problems in Section 2.3 and prove them in Section 5. We consider a new complexity class, $\mathsf{BC_{=}P}$, in Section 6, and we list some open problems in Section 7.

2.1 Exact completeness results

For a joint distribution D over $(\{0,1\}^k)^n$, we say that D_i, D_j are *uncorrelated* if they have covariance 0, in other words $\mathrm{E}(D_i \cdot D_j) = \mathrm{E}(D_i) \cdot \mathrm{E}(D_j)$ (when $\{0,1\}^k$ is interpreted as binary representations of integers from 0 to $2^k - 1$). Uncorrelated is the same as independent if $k = 1$. We consider the following extreme notion of a distribution being nonuniform.

Definition 2.1. A joint distribution is *discordant* if there are ≥ 2 coordinates and every pair of coordinates is neither uncorrelated nor identically distributed.

Definition 2.2. PANOPTIC-STATS is the following promise problem.

$$\begin{aligned} \text{PANOPTIC-STATS}_{\text{YES}} &= \big\{S \,:\, S(U) \text{ is uniform}\big\}, \\ \text{PANOPTIC-STATS}_{\text{NO}} &= \big\{S \,:\, S(U) \text{ is discordant}\big\}. \end{aligned}$$

We say that promise problem Π is a generalization of promise problem Π', or that Π' is a restriction of Π, if $\Pi'_{\text{YES}} \subseteq \Pi_{\text{YES}}$ and $\Pi'_{\text{NO}} \subseteq \Pi_{\text{NO}}$.

Fact 2.3. PANOPTIC-STATS *is generalized by all the following languages, which are defined in a natural way.*

> UNIFORM, IID, FULLY-INDEPENDENT, IDENTICALLY-DISTRIBUTED, EXCHANGEABLE, K-WISE-UNIFORM, K-WISE-INDEPENDENT, K-WISE-EXCHANGEABLE, 2-WISE-UNCORRELATED, K-EXISTS-UNIFORM, K-EXISTS-INDEPENDENT, K-EXISTS-IDENTICALLY-DISTRIBUTED, K-EXISTS-EXCHANGEABLE, 2-EXISTS-UNCORRELATED, NON-DISCORDANT.

For example, $S \in$ UNIFORM $\iff$ $S(U)$ is uniform. Also, $K \geq 2$ is any constant (unrelated to k). Technical caveat: To ensure the K-WISE- and K-EXISTS- problems generalize PANOPTIC-STATS, *they are defined in terms of a property holding for every or some (respectively) set of* $\min(K,n)$ *coordinates.*

We prove that PANOPTIC-STATS and all the languages listed in Fact 2.3 are complete for the complexity class $\mathsf{C_{=}P}$ (defined below). In fact, the $\mathsf{C_{=}P}$-hardness of each of the individual languages in Fact 2.3 is fairly simple to prove, but the $\mathsf{C_{=}P}$-hardness of PANOPTIC-STATS shows two things: (1) that this phenomenon is very robust, not dependent on some fragile aspects of the properties being decided, and (2) that only one proof is needed to show the $\mathsf{C_{=}P}$-hardness of all the languages in Fact 2.3.

To prove the $\mathsf{C_{=}P}$-hardness of PANOPTIC-STATS, it suffices to prove hardness for the case $n = 2$. However, hardness for $n = 2$ does not seem to directly imply hardness for a larger number of coordinates;

it is desirable to prove hardness even when restricted to samplers that are small in terms of the number of coordinates n. We formalize this by introducing a new parameter m and viewing k,n,s as functions of m. Thus m can be thought of as indexing a family of parameter settings.

Definition 2.4. We say that a triple of functions $\kappa(m),\nu(m),\sigma(m):\mathbb{N}\to\mathbb{N}$ is *polite* if the functions are monotonically nondecreasing, polynomially bounded in m, computable in time polynomial in m, and $\sigma(m)\geq m$.

Definition 2.5. PANOPTIC-STATS$^{\kappa,\nu,\sigma}$ is the restriction of PANOPTIC-STATS to (k,n,s)-samplers with $k=\kappa(m)$, $n=\nu(m)$, and $s\leq\sigma(m)$ for some m, where κ,ν,σ is assumed to be polite.

We now state the definition of our central complexity class, $\mathsf{C_{=}P}$. We use a standard model of computation in which randomized algorithms have access to independent unbiased coin flips.

Definition 2.6. $\mathsf{prC_{=}P}$ is the class of all promise problems for which there exists a polynomial-time randomized algorithm M that accepts with probability $1/2$ on YES instances, and accepts with probability $\neq 1/2$ on NO instances. Also, $\mathsf{C_{=}P}$ is defined as the class of languages in $\mathsf{prC_{=}P}$.

Proposition 2.7. $\mathsf{prC_{=}P}$ *is the class of all promise problems Karp-reducible to the following promise problem,* UNIFORM-BIT.

$$\begin{aligned}\text{UNIFORM-BIT}_{\text{YES}} &= \{S : S \textit{ is a } (1,1)\textit{-sampler and } S(U) \textit{ is uniform}\},\\ \text{UNIFORM-BIT}_{\text{NO}} &= \{S : S \textit{ is a } (1,1)\textit{-sampler and } S(U) \textit{ is nonuniform}\}.\end{aligned}$$

Proof. Suppose $\Pi\in\mathsf{prC_{=}P}$ is witnessed by M taking an input x and a uniformly random string y of some polynomial length. To reduce Π to UNIFORM-BIT, map x to S_x where $S_x(y)=M(x,y)$. Conversely, suppose Π reduces to UNIFORM-BIT. Then $\Pi\in\mathsf{prC_{=}P}$ is witnessed by M that takes x, runs the reduction to get a $(1,1)$-sampler S_x, and runs S_x on a uniformly random input. ☐

Theorem 2.8. PANOPTIC-STATS$^{\kappa,\nu,\sigma}$ *is* $\mathsf{prC_{=}P}$*-hard for every polite* κ,ν,σ *with* $\kappa\nu\leq o(\sigma)$.

Theorem 2.9. PANOPTIC-STATS$^{\kappa,\nu,\sigma}$ *is* $\mathsf{prC_{=}P}$*-hard even when restricted to samplers that are* AC^0*-type circuits with depth* 3 *and top fan-in* 2*, for every polite* κ,ν,σ *with* $\kappa\nu+\nu^2\leq o(\sigma)$.

Theorem 2.10. *All the languages listed in Fact 2.3 are in* $\mathsf{C_{=}P}$.

Consequently, all the languages listed in Fact 2.3 are $\mathsf{C_{=}P}$-complete, even when restricted to (κ,ν,σ)-samplers (like in Definition 2.5) with polite κ,ν,σ satisfying $\kappa\nu\leq o(\sigma)$ (for general circuit samplers) or satisfying $\kappa\nu+\nu^2\leq o(\sigma)$ (for depth-3 circuits with top fan-in 2).

We mention that some problems concerning conditional independence are also $\mathsf{C_{=}P}$-complete. For example, deciding whether the first $n-1$ coordinates of $S(U)$ are fully independent conditioned on the last coordinate is at least as hard as the corresponding non-conditional problem. Another problem concerning conditional independence is whether $S(U)$ forms a (time-inhomogeneous) Markov chain (assuming $n\geq 3$). The construction in our proof of Theorem 2.8 also shows that this problem is $\mathsf{C_{=}P}$-hard. Both these problems are in $\mathsf{C_{=}P}$ by the same techniques used in the proof of Theorem 2.10.

2.2 Exact algorithmic results

We say a (k,n,s)-sampler is *d-local* if each of the kn output bits depends on at most d of the uniformly random input bits. For d-local samplers, if $dk \leq O(\log s)$ then some statistical properties, such as being pairwise independent or having identically distributed marginals, can be decided trivially in polynomial time. We now prove that some other properties, namely being fully independent or being exchangeable, can be decided in polynomial time when $d = 2$ and $k = 1$. (Admittedly, our algorithms are not very "algorithmic"; we prove combinatorial characterizations for which it is trivial to check whether a given sampler satisfies the characterization.)

Theorem 2.11. *There exists a linear-time algorithm for deciding whether the joint distribution of a given 2-local $(1,n)$-sampler is fully independent.*

Theorem 2.12. *There exists a linear-time algorithm for deciding whether the joint distribution of a given 2-local $(1,n)$-sampler is exchangeable.*

When $d = 2$ and $k = 1$, we can also improve the efficiency of the trivial quadratic-time algorithm for deciding pairwise independence.

Theorem 2.13. *There exists a linear-time Karp reduction from the problem of deciding whether the joint distribution of a given 2-local $(1,n)$-sampler is pairwise independent, to the element distinctness problem. Hence the former problem can be solved in deterministic $O(n \log n)$ time and in zero-error randomized expected linear time.*

One can also consider logspace samplers that have streaming/one-way access to their random input bits, and which are usually modeled as layered read-once branching programs representing a certain type of (time-inhomogeneous) Markov chain. For logspace samplers, some statistical properties, such as being pairwise independent or having identically distributed marginals, can be decided in polynomial time by straightforward dynamic programming algorithms; the complexities of deciding full independence and exchangeability remain open.

2.3 Approximate completeness results

We quantify approximation in terms of statistical distance (also known as total variation distance).

Definition 2.14. The statistical distance between two distributions $D^{(1)}, D^{(2)}$ over the same set is defined as

$$\begin{aligned}\left\|D^{(1)} - D^{(2)}\right\| &= \max_{\text{events } E} \left|\Pr[D^{(1)} \in E] - \Pr[D^{(2)} \in E]\right| \\ &= \max_{\text{events } E} \left(\Pr[D^{(1)} \in E] - \Pr[D^{(2)} \in E]\right) \\ &= \frac{1}{2} \cdot \sum_{\text{outcomes } w} \left|\Pr[D^{(1)} = w] - \Pr[D^{(2)} = w]\right| .\end{aligned}$$

We say $D^{(1)}, D^{(2)}$ are c-close if $\left\|D^{(1)} - D^{(2)}\right\| \leq c$, and f-far if $\left\|D^{(1)} - D^{(2)}\right\| \geq f$.

We prove that for appropriate parameters, approximate versions of the full independence and exchangeability problems are prSZK-complete (for arbitrary circuit samplers). We do not reproduce the original definition of prSZK, but we make use of the characterization of this class proved by Sahai and Vadhan [31]. The following is our general formulation of the approximate full independence problem.

Definition 2.15. For functions $0 \le c(k,n,s) < f(k,n,s) \le 1$, FULLY-INDEPENDENT^{c,f} is the following promise problem.[3]

$$\begin{aligned}
\text{FULLY-INDEPENDENT}^{c,f}_{\text{YES}} &= \big\{S \,:\, S \text{ is a } (k,n,s)\text{-sampler and } S(U) \text{ is } c(k,n,s)\text{-close} \\
&\qquad \text{to some fully independent distribution over } (\{0,1\}^k)^n\big\}\,, \\
\text{FULLY-INDEPENDENT}^{c,f}_{\text{NO}} &= \big\{S \,:\, S \text{ is a } (k,n,s)\text{-sampler and } S(U) \text{ is } f(k,n,s)\text{-far} \\
&\qquad \text{from every fully independent distribution over } (\{0,1\}^k)^n\big\}\,.
\end{aligned}$$

Theorem 2.16. FULLY-INDEPENDENT^{c,f} *is* prSZK-*hard for all constants* $0 < c < f < 1/4$.

Theorem 2.17. FULLY-INDEPENDENT$^{c,f} \in$ prSZK *where* $c = c'/(n+1)$, *for all constants* $0 < c' < f^2 < 1$.

We have, for example, that FULLY-INDEPENDENT$^{0.05/(n+1),\ 0.24}$ is prSZK-complete. The containment in prSZK follows from Theorem 2.17. Although the prSZK-hardness does not follow from the statement of Theorem 2.16, the proof indeed yields this; we stated Theorem 2.16 using constants for the sake of simplicity and clarity. (It is open to prove Theorem 2.17 with constant c.)

Definition 2.18. For functions $0 \le c(k,n,s) < f(k,n,s) \le 1$, EXCHANGEABLE^{c,f} is the following promise problem.

$$\begin{aligned}
\text{EXCHANGEABLE}^{c,f}_{\text{YES}} &= \big\{S \,:\, S \text{ is a } (k,n,s)\text{-sampler and } S(U) \text{ is } c(k,n,s)\text{-close} \\
&\qquad \text{to some exchangeable distribution over } (\{0,1\}^k)^n\big\}\,, \\
\text{EXCHANGEABLE}^{c,f}_{\text{NO}} &= \big\{S \,:\, S \text{ is a } (k,n,s)\text{-sampler and } S(U) \text{ is } f(k,n,s)\text{-far} \\
&\qquad \text{from every exchangeable distribution over } (\{0,1\}^k)^n\big\}\,.
\end{aligned}$$

Theorem 2.19. EXCHANGEABLE^{c,f} *is* prSZK-*hard for all constants* $0 < c < f < 1/2$.

Theorem 2.20. EXCHANGEABLE$^{c,f} \in$ prSZK *for all constants* $0 < 2c < f^2 < 1$.

Consequently, for example, EXCHANGEABLE$^{0.12,\ 0.49}$ is prSZK-complete.

3 Proofs of exact completeness results

We prove a key lemma in Section 3.1. Then we use the key lemma to prove Theorem 2.8 and Theorem 2.9 in Section 3.2. Then we prove Theorem 2.10 in Section 3.3.

[3]The superscripts have a different meaning than the superscripts in Definition 2.5.

3.1 The key lemma

The following is the key lemma in the proof of Theorem 2.8. It can be interpreted qualitatively as a certain type of amplification.

Lemma 3.1. *There is an algorithm that takes as input a $(1,1,s)$-sampler S and an integer $n \geq 2$, runs in time $O(n+s)$, and outputs a $(1,n,O(n+s))$-sampler T such that the following both hold.*

$$\begin{aligned} S(U) \text{ is uniform} &\implies T(U) \text{ is uniform}, \\ S(U) \text{ is nonuniform} &\implies T(U) \text{ is discordant}. \end{aligned}$$

Proof. Let T perform the following computation.

run S and let b be its output
choose bits $a_1, a_2, \ldots, a_n$ uniformly at random
if there exists an $\ell < n$ such that $a_\ell = 0$ **then**
  let ℓ^* be the least such ℓ
  output $a_1, \ldots, a_{\ell^*}, b, a_{\ell^*+2}, \ldots, a_n$
else output $a_1, \ldots, a_n$

It is straightforward to see that if $S(U)$ is uniform then $T(U)$ is uniform. Now suppose $S(U)$ is nonuniform, say $\Pr[S(U)=1] = p \neq 1/2$. For brevity we define $D = T(U)$. Consider any two coordinates D_i and D_j where $i < j$. For technical reasons in the analysis below, if ℓ^* does not exist then we define ℓ^* to be an arbitrary value $> n$.

We first show that D_i and D_j are not identically distributed. If $i > 1$ then

$$\begin{aligned} \Pr[D_i = 1] \;&=\; \Pr\big[D_i = 1 \;\big|\; \ell^* = i-1\big] \cdot \Pr[\ell^* = i-1] \;+\; \Pr\big[D_i = 1 \;\big|\; \ell^* \neq i-1\big] \cdot \Pr[\ell^* \neq i-1] \\ &=\; p \cdot \frac{1}{2^{i-1}} \;+\; \frac{1}{2} \cdot \left(1 - \frac{1}{2^{i-1}}\right). \end{aligned}$$

Similarly,

$$\Pr[D_j = 1] = p \cdot \frac{1}{2^{j-1}} + \frac{1}{2} \cdot \left(1 - \frac{1}{2^{j-1}}\right).$$

Since $p \neq 1/2$, and since $\Pr[D_i = 1]$ and $\Pr[D_j = 1]$ are different convex combinations of p and $1/2$, that means they are not equal. More formally,

$$\Pr[D_i = 1] - \Pr[D_j = 1] \;=\; \left(p - \frac{1}{2}\right)\left(\frac{1}{2^{i-1}} - \frac{1}{2^{j-1}}\right) \;\neq\; 0\,.$$

On the other hand, suppose $i = 1$. Then $\Pr[D_i = 1] = 1/2$, and $\Pr[D_j = 1]$ is a nontrivial convex combination of p and $1/2$ and is thus not equal to $\Pr[D_i = 1]$. In either case, D_i and D_j are not identically distributed.

Now we show that D_i and D_j are correlated. Suppose $j = i+1$. Then $\Pr\big[D_j = 1 \,\big|\, D_i = 1\big] = 1/2$, and

$$\begin{aligned}\Pr\big[D_j = 1 \,\big|\, D_i = 0\big] &= \Pr\big[D_j = 1 \,\big|\, \ell^* = i,\, D_i = 0\big] \cdot \Pr\big[\ell^* = i \,\big|\, D_i = 0\big] + \\ &\quad \Pr\big[D_j = 1 \,\big|\, \ell^* < i,\, D_i = 0\big] \cdot \Pr\big[\ell^* < i \,\big|\, D_i = 0\big] \\ &= p \cdot \Pr\big[\ell^* = i \,\big|\, D_i = 0\big] + \frac{1}{2} \cdot \big(1 - \Pr\big[\ell^* = i \,\big|\, D_i = 0\big]\big).\end{aligned}$$

(Technically $\Pr\big[D_j = 1 \,\big|\, \ell^* < i,\, D_i = 0\big]$ is undefined if $i = 1$, but then $1 - \Pr\big[\ell^* = i \,\big|\, D_i = 0\big] = 0$ anyway so the final equation above still holds.) It follows that

$$\Pr\big[D_j = 1 \,\big|\, D_i = 0\big] - \Pr\big[D_j = 1 \,\big|\, D_i = 1\big] \;=\; \Big(p - \frac{1}{2}\Big) \cdot \Pr\big[\ell^* = i \,\big|\, D_i = 0\big] \;\neq\; 0$$

since $p \neq 1/2$ and $\Pr\big[\ell^* = i \,\big|\, D_i = 0\big] > 0$. On the other hand, suppose $j > i+1$. Then

$$\Pr\big[D_j = 1 \,\big|\, D_i = 0\big] = 1/2\,,$$

and

$$\begin{aligned}\Pr\big[D_j = 1 \,\big|\, D_i = 1\big] &= \Pr\big[D_j = 1 \,\big|\, \ell^* = j-1,\, D_i = 1\big] \cdot \Pr\big[\ell^* = j-1 \,\big|\, D_i = 1\big] + \\ &\quad \Pr\big[D_j = 1 \,\big|\, \ell^* \neq j-1,\, D_i = 1\big] \cdot \Pr\big[\ell^* \neq j-1 \,\big|\, D_i = 1\big] \\ &= p \cdot \Pr\big[\ell^* = j-1 \,\big|\, D_i = 1\big] + \frac{1}{2} \cdot \big(1 - \Pr\big[\ell^* = j-1 \,\big|\, D_i = 1\big]\big).\end{aligned}$$

It follows that

$$\Pr\big[D_j = 1 \,\big|\, D_i = 1\big] - \Pr\big[D_j = 1 \,\big|\, D_i = 0\big] \;=\; \Big(p - \frac{1}{2}\Big) \cdot \Pr\big[\ell^* = j-1 \,\big|\, D_i = 1\big] \;\neq\; 0$$

since $p \neq 1/2$ and $\Pr\big[\ell^* = j-1 \,\big|\, D_i = 1\big] > 0$. In either case, D_i and D_j are correlated since

$$\Pr\big[D_j = 1 \,\big|\, D_i = 0\big] \neq \Pr\big[D_j = 1 \,\big|\, D_i = 1\big]\,. \qquad \square$$

Lemma 3.2. *Lemma 3.1 holds even when T is required to be an AC^0-type circuit with depth 3 and top fan-in 2, except that the size of T and the running time of the algorithm both become $O(n^2 + s)$.*

Proof. The construction and analysis are the same as in the proof of Lemma 3.1, but we need more care in implementing T. First, we use a standard reduction to convert S into a 3-CNF F that accepts the same number of inputs as S (but has more input bits). Thus, for some polynomially large q, S accepts a uniformly random input with probability $1/2$ iff F accepts a uniformly random input with probability $1/2^q$. Let $x_1, x_2, \dots, x_r$ denote the input bits of F. Construct a new CNF F' with input bits $x_0, x_1, \dots, x_r$ by taking F and including $\overline{x_0}$ in each of the clauses (yielding a 4-CNF), then adding a new clause $(x_0 \vee x_1 \vee \cdots \vee x_q)$. Since

$$\Pr[F' \text{ accepts}] \;=\; \frac{1}{2} \cdot \Pr[F \text{ accepts}] + \frac{1}{2} \cdot \Pr\big[(x_1 \vee \cdots \vee x_q) \text{ accepts}\big]$$

it follows that F accepts with probability $1/2^q$ iff F' accepts with probability $1/2$. Now to implement T, we include a copy of F' as well as the random input bits $a_1, a_2, \dots, a_n$. The 1^{st} output bit of T is just a_1. For the i^{th} output bit when $i > 1$, we have a multiplexer that selects the output of F' if $(a_1 \wedge a_2 \wedge \cdots \wedge a_{i-2} \wedge \overline{a_{i-1}})$ is true, and selects a_i otherwise. Overall, T is an OR-AND-OR circuit (with negations pushed to the inputs) where each output gate has fan-in at most 2. $\square$

3.2 $\mathsf{prC_=P}$-hardness

We need one final corollary before we are ready to put the pieces together to prove Theorem 2.8 and Theorem 2.9.

Corollary 3.3. *Lemma 3.1 and Lemma 3.2 also hold when the algorithm is additionally given an integer $k \geq 1$ and is required to output a (k,n)-sampler T, except that the size of T and the running time of the algorithm both become $O(kn+s)$ (for Lemma 3.1) or $O(kn+n^2+s)$ (for Lemma 3.2).*

Proof. If T' is the output of the algorithm from Lemma 3.1 or Lemma 3.2, we can trivially modify it into a sampler T that prepends independent uniformly random bit strings of length $k-1$ to the n coordinates. In the YES case, $T(U)$ is still uniform. Consider the NO case. The property that no two coordinates are identically distributed is inherited from T'. To see that coordinates $T(U)_i, T(U)_j$ are still correlated, abbreviate $T(U)$ as D and $T'(U)$ as D', and let $D_i = D'_i + I$ and $D_j = D'_j + J$ where I, J are independent uniformly random even numbers in the range $\{0,\ldots,2^k-2\}$, and note that

$$\begin{aligned}
\mathrm{E}(D_iD_j) &= \mathrm{E}(D'_iD'_j)+\mathrm{E}(D'_iJ)+\mathrm{E}(ID'_j)+\mathrm{E}(IJ)\\
&= \mathrm{E}(D'_iD'_j)+\mathrm{E}(D'_i)\,\mathrm{E}(J)+\mathrm{E}(I)\,\mathrm{E}(D'_j)+\mathrm{E}(I)\,\mathrm{E}(J)\\
&\neq \mathrm{E}(D'_i)\,\mathrm{E}(D'_j)+\mathrm{E}(D'_i)\,\mathrm{E}(J)+\mathrm{E}(I)\,\mathrm{E}(D'_j)+\mathrm{E}(I)\,\mathrm{E}(J)\\
&= \big(\mathrm{E}(D'_i)+\mathrm{E}(I)\big)\big(\mathrm{E}(D'_j)+\mathrm{E}(J)\big)\\
&= \mathrm{E}(D_i)\,\mathrm{E}(D_j)\,.
\end{aligned}$$

□

Proof of Theorem 2.8. We reduce UNIFORM-BIT to PANOPTIC-STATS$^{\kappa,\nu,\sigma}$. Let c be the constant factor in the big O in Corollary 3.3. Given a $(1,1,s)$-sampler S, we first find the smallest m such that $c \cdot \big(\kappa(m)\nu(m)+s\big) \leq \sigma(m)$. Such an m exists and is $O(s)$ because $\kappa\nu \leq o(\sigma)$ and $\sigma(m) \geq m$ for all m. Then we run the algorithm from Corollary 3.3 (based on Lemma 3.1) with $k=\kappa(m)$ and $n=\nu(m)$ to get T of size at most $c\cdot\big(\kappa(m)\nu(m)+s\big) \leq \sigma(m)$. Thus the following both hold.

$$\begin{aligned}
S \in \text{UNIFORM-BIT}_{\text{YES}} &\implies T \in \text{PANOPTIC-STATS}^{\kappa,\nu,\sigma}_{\text{YES}}\,,\\
S \in \text{UNIFORM-BIT}_{\text{NO}} &\implies T \in \text{PANOPTIC-STATS}^{\kappa,\nu,\sigma}_{\text{NO}}\,.
\end{aligned}$$

The reduction's running time is polynomial since $m, \kappa(m), \nu(m), \sigma(m)$ are all polynomially bounded in s and computable in time polynomial in s, and since the algorithm from Corollary 3.3 runs in time $O(kn+s)$. □

Proof of Theorem 2.9. We reduce UNIFORM-BIT to PANOPTIC-STATS$^{\kappa,\nu,\sigma}$ restricted as in the statement of Theorem 2.9. Let c be the constant factor in the big O in Corollary 3.3. Given a $(1,1,s)$-sampler S, we first find the smallest m such that $c\cdot\big(\kappa(m)\nu(m)+\nu(m)^2+s\big) \leq \sigma(m)$. Such an m exists and is $O(s)$ because $\kappa\nu+\nu^2 \leq o(\sigma)$ and $\sigma(m) \geq m$ for all m. Then we run the algorithm from Corollary 3.3 (based on Lemma 3.2) with $k=\kappa(m)$ and $n=\nu(m)$ to get a depth-3, top fan-in 2 circuit T of size at most $c\cdot\big(\kappa(m)\nu(m)+\nu(m)^2+s\big) \leq \sigma(m)$. Thus the following both hold.

$$\begin{aligned}
S \in \text{UNIFORM-BIT}_{\text{YES}} &\implies T \in \text{PANOPTIC-STATS}^{\kappa,\nu,\sigma}_{\text{YES}}\,,\\
S \in \text{UNIFORM-BIT}_{\text{NO}} &\implies T \in \text{PANOPTIC-STATS}^{\kappa,\nu,\sigma}_{\text{NO}}\,.
\end{aligned}$$

The reduction's running time is polynomial since $m, \kappa(m), \nu(m), \sigma(m)$ are all polynomially bounded in s and computable in time polynomial in s, and since the algorithm from Corollary 3.3 runs in time $O(kn+n^2+s)$. $\square$

3.3 Containment in $\mathsf{C_{=}P}$

In the proof of Theorem 2.10 we use the following lemma, which states that $\mathsf{C_{=}P}$ is closed under exponential conjunctions and polynomial disjunctions. We supply a folklore proof of this lemma in Section A.1.

Lemma 3.4. *If $L \in \mathsf{C_{=}P}$ then both of the following hold.*

- $\forall_q L \in \mathsf{C_{=}P}$ *for every polynomial q, where* $\forall_q L = \{x : (x,y) \in L \text{ for all } y \in \{0,1\}^{q(|x|)}\}$.
- $\vee L \in \mathsf{C_{=}P}$ *where* $\vee L = \{(x_1,\dots,x_\ell) : x_i \in L \text{ for some } i\}$.

Proof of Theorem 2.10. The arguments are very similar, so we just give three representative examples: FULLY-INDEPENDENT, K-WISE-EXCHANGEABLE, and 2-EXISTS-UNCORRELATED. First we mention a useful tool: If S_1, S_2 are $(1,1)$-samplers, then we define $\mathrm{Equ}(S_1,S_2)$ to be a $(1,1)$-sampler that picks $i \in \{1,2\}$ uniformly at random, runs S_i, and negates the output if $i=2$. Hence $\mathrm{Equ}(S_1,S_2)(U)$ is uniform iff $S_1(U), S_2(U)$ are identically distributed.

Now we prove that FULLY-INDEPENDENT $\in \mathsf{C_{=}P}$. Note that FULLY-INDEPENDENT $= \forall_q L$ where, if we view S as (say) a (k,n)-sampler, and y as (an appropriately encoded description of) an element of $(\{0,1\}^k)^n$ (so q is linear in the size of S), then

$$(S,y) \in L \iff \Pr[S(U)=y] = \textstyle\prod_{i=1}^n \Pr[S(U)_i = y_i]\,.$$

Thus by Lemma 3.4 it suffices to show that $L \in \mathsf{C_{=}P}$. A reduction from L to UNIFORM-BIT just outputs $\mathrm{Equ}(S_1,S_2)$, where S_1 runs S and accepts iff the output is y, and S_2 runs S for n times and accepts iff for all i, the i^{th} coordinate of the output of the i^{th} run is y_i.

Now we prove that K-WISE-EXCHANGEABLE $\in \mathsf{C_{=}P}$. Note that K-WISE-EXCHANGEABLE $= \forall_q L$ where, if we view S as (say) a (k,n)-sampler, and $y = (I,\pi,w)$ as (an appropriately encoded description of) a subset $I \subseteq \{1,\dots,n\}$ of size $\min(K,n)$, a permutation π on $\{1,\dots,\min(K,n)\}$, and an element $w \in (\{0,1\}^k)^{\min(K,n)}$ (so q is certainly polynomial in the size of S), then

$$\big(S,(I,\pi,w)\big) \in L \iff \Pr[S(U)_I = w] = \Pr[S(U)_I = \pi(w)]$$

where $S(U)_I$ is the restriction to coordinates indexed by I, and $\pi(w) \in (\{0,1\}^k)^{\min(K,n)}$ is obtained by permuting the coordinates of w by π. Thus by Lemma 3.4 it suffices to show that $L \in \mathsf{C_{=}P}$. A reduction from L to UNIFORM-BIT just outputs $\mathrm{Equ}(S_1,S_2)$, where S_1 runs S and accepts iff the output restricted to I is w, and S_2 runs S and accepts iff the output restricted to I is $\pi(w)$.

Now we prove that 2-EXISTS-UNCORRELATED $\in \mathsf{C_{=}P}$. Note that if we define the language

$$L = \big\{(S,i,j) : S(U)_i \text{ and } S(U)_j \text{ are uncorrelated}\big\}\,,$$

then 2-EXISTS-UNCORRELATED reduces to $\vee L$ by mapping a (k,n)-sampler S to

$$\big((S,1,2),(S,1,3),(S,1,4),\dots,(S,n-1,n)\big).$$

Thus by Lemma 3.4 it suffices to show that $L \in \mathsf{C_{=}P}$. A reduction from L to UNIFORM-BIT just outputs $\mathrm{Equ}(S_1,S_2)$, where S_1 runs S yielding some $y \in (\{0,1\}^k)^n$ and accepts with probability $(1/2^{2k})\cdot y_i \cdot y_j$ so that

$$\Pr[S_1(U)=1] \;=\; \frac{1}{2^{2k}}\cdot \mathrm{E}\big(S(U)_i \cdot S(U)_j\big),$$

and S_2 runs S twice (independently) yielding some $y^{(1)}$ and $y^{(2)}$ and accepts with probability $(1/2^{2k})\cdot y_i^{(1)} \cdot y_j^{(2)}$ so that

$$\Pr[S_2(U)=1] \;=\; \frac{1}{2^{2k}}\cdot \mathrm{E}(S(U)_i)\cdot \mathrm{E}(S(U)_j). \qquad \square$$

4 Proofs of exact algorithmic results

We prove Theorem 2.11, Theorem 2.12, and Theorem 2.13 in Section 4.1, Section 4.2, and Section 4.3, respectively.

First we introduce some terminology to describe 2-local samplers. Each output bit depends on either zero, one, or two input bits. Output bits that depend on zero input bits are constants (0 or 1). The nonconstant output bits can be modeled with an undirected graph (multi-edges and self-loops allowed) as follows. The input bits are indexed by the nodes. Each output bit depending on one input bit is a self-loop, labeled with a function from $\{0,1\}$ to $\{0,1\}$ (either the identity or negation). Each output bit depending on two input bits is an edge between those two nodes, labeled with a function from $\{0,1\}^2$ to $\{0,1\}$. There are three types of such functions that depend on both bits: AND-type (accepting one of the four inputs), XOR-type (accepting two of the four inputs), and OR-type (accepting three of the four inputs).

4.1 Full independence for 2-local samplers

We prove Theorem 2.11. Consider a 2-local $(1,n)$-sampler S, and assume without loss of generality that S has no constant output bits. We claim that $S(U)$ is fully independent iff both of the following conditions hold.

(i) The graph is a forest, ignoring self-loops.

(ii) Each connected component of the graph has at most one of the following: a self-loop, an AND-type edge, or an OR-type edge.

It is trivial to check in linear time whether these conditions hold.

First we assume that (i) and (ii) both hold, and show that $S(U)$ is fully independent. The different connected components of the graph are certainly fully independent of each other, so we can focus on showing that the coordinates of a single connected component are fully independent. If there is a self-loop, an AND-type edge, or an OR-type edge in the connected component, then let e be that edge. Otherwise, let e be any edge in the connected component. We show that conditioned on e evaluating to any particular bit, the joint distribution of the remaining edges in e's connected component is uniform. This implies that the whole joint distribution of the connected component is fully independent.

Suppose e is a self-loop at some node v, so we are conditioning on v being some particular bit. Ignoring e itself, we can view e's connected component as a tree rooted at v with only XOR-type edges. After the conditioning, there is a bijection between the set of all assignments of values to the edges (excluding e) and the set of all assignments of values to the nodes (excluding v) in e's connected component: An assignment to nodes (together with the conditioned value of v) determines an assignment to edges. Furthermore, every assignment to edges arises from some assignment to nodes, because for any assignment to edges, we can start at v and work our way downward to the leaves, uniquely specifying the value of each node in terms of the values of its parent and the edge to its parent. Since the sets have the same size, we have exhibited a bijection between them. This means that conditioned on either value of e, the joint distribution of all the other edges in e's connected component is uniform.

Now suppose $e = \{u, v\}$ is not a self-loop. We show that, in fact, conditioned on any one of the four assignments of values to the pair u, v, the joint distribution of all the other edges in e's connected component is uniform. Removing e results in two new connected components, each of which is a tree of XOR-type edges, one rooted at u and the other rooted at v. Let U denote the set of nodes in u's new connected component excluding u itself, and let V denote the set of nodes in v's new connected component excluding v itself. By the argument from the previous paragraph (when e was a self-loop), a uniformly random assignment to U induces a uniformly random assignment to the edges in u's new connected component, and similarly for V. Since assignments to U and V are chosen independently of each other, this means that the values of all the edges in e's original connected component (except e itself) are jointly uniformly distributed (conditioned on any particular assignment to u, v, and hence conditioned on any particular assignment to e).

Now we prove the converse by assuming that (i) and (ii) do not both hold, and showing that $S(U)$ is not fully independent. Let us refer to self-loops, AND-type edges, and OR-type edges as *non-XOR-type* edges. If (i) and (ii) do not both hold, then at least one of the following conditions holds.

(A) There is a cycle consisting entirely of XOR-type edges.

(B) There is a cycle with exactly one AND-type edge or OR-type edge.

(C) There is a path between two non-XOR-type edges.

Suppose (A) holds. Let e be an edge on the cycle. Then e's marginal distribution is uniform, but conditioning on any particular values of the other edges on the cycle determines whether or not e's endpoints are the same bit as each other, and thus fixes the value of e. Hence $S(U)$ is not fully independent. Suppose (B) holds. Let ℓ denote the number of nodes on the cycle. Then the probability that all edges on the cycle evaluate to 1 must be an integer multiple of $1/2^{\ell}$ (since they only depend on ℓ input bits), but the product of the marginal probabilities that each edge on the cycle evaluates to 1 must be either $1/2^{\ell+1}$ (if there is an AND-type edge) or $3/2^{\ell+1}$ (if there is an OR-type edge). Hence $S(U)$ is not fully independent. Suppose (C) holds. Without loss of generality, all intermediate edges on the path are XOR-type. Let e_1 and e_2 be the two non-XOR-type edges, which we consider to be part of the path. Let ℓ denote the number of nodes on the path. Then the probability that all edges on the path evaluate to 1 must be an integer multiple of $1/2^{\ell}$ (since they only depend on ℓ input bits), but the product of the marginal probabilities that each edge on the path evaluates to 1 must be either $1/2^{\ell+1}$ (if neither e_1 nor e_2 is OR-type) or $3/2^{\ell+1}$ (if exactly one of e_1, e_2 is OR-type) or $9/2^{\ell+1}$ (if both e_1 and e_2 are OR-type). Hence $S(U)$ is not fully independent.

4.2 Exchangeability for 2-local samplers

We prove Theorem 2.12. We begin with a lemma.

Lemma 4.1. *A joint distribution D over $(\{0,1\}^1)^n$ is exchangeable iff both of the following conditions hold.*

(1) The marginals D_i are all identically distributed.

(2) For all $i \neq j$, if $\Pr[D_i \neq D_j] > 0$ then the joint distribution of the other $n-2$ coordinates is the same when conditioned on $(D_i = 1, D_j = 0)$ as it is when conditioned on $(D_i = 0, D_j = 1)$.

Proof of Lemma 4.1. Suppose D is exchangeable. Then (1) holds trivially. To see that (2) holds, consider $i \neq j$ such that $\Pr[D_i \neq D_j] > 0$. First note that since D_i, D_j are identically distributed,

$$\Pr[D_i = 1, D_j = 0] = \Pr[D_i = 0, D_j = 1] > 0\,.$$

For some arbitrary particular bits b_h (for $h \notin \{i, j\}$), let E denote the event that $D_h = b_h$ for all $h \notin \{i, j\}$. Then we have

$$\begin{aligned}
\Pr\big[E \mid D_i = 1, D_j = 0\big] &= \frac{\Pr\big[E \text{ and } D_i = 1, D_j = 0\big]}{\Pr[D_i = 1, D_j = 0]} \\
&= \frac{\Pr\big[E \text{ and } D_i = 0, D_j = 1\big]}{\Pr[D_i = 0, D_j = 1]} \\
&= \Pr\big[E \mid D_i = 0, D_j = 1\big]\,.
\end{aligned}$$

where

$$\Pr\big[E \text{ and } D_i = 1, D_j = 0\big] = \Pr\big[E \text{ and } D_i = 0, D_j = 1\big]$$

holds by exchangeability. This shows that (2) holds.

For the converse, suppose (1) and (2) both hold. Since every permutation is a composition of transpositions, it suffices to show that the joint distribution is invariant under transposing coordinates. Let D' be obtained from D by transposing some coordinates $i \neq j$. For some arbitrary particular bits b_h (for $h \in \{1, \dots, n\}$), let E denote the event that $D_h = b_h$ for all $h \in \{1, \dots, n\}$, and let E' denote the event that $D'_h = b_h$ for all $h \in \{1, \dots, n\}$. To show that D and D' are equal as distributions, we just need to show that $\Pr[E] = \Pr[E']$. If $b_i = b_j$ then this certainly holds since E and E' are the same event. If $b_i \neq b_j$ and $\Pr[D_i \neq D_j] = 0$ then $\Pr[E] = \Pr[E'] = 0$. Finally, suppose $b_i \neq b_j$ and $\Pr[D_i \neq D_j] > 0$. Assume $b_i = 1$ and $b_j = 0$; the other case is symmetric. Since D_i, D_j are identically distributed by (1), it follows that

$$\Pr[D_i = 1, D_j = 0] = \Pr[D_i = 0, D_j = 1] = \Pr[D'_i = 1, D'_j = 0] > 0\,.$$

Also note that

$$\Pr\big[E \;\big|\; D_i = 1, D_j = 0\big] \;=\; \Pr\big[E' \;\big|\; D'_i = 1, D'_j = 0\big]$$

by (2) and the definition of D'. Putting the pieces together, we have

$$\begin{aligned}\Pr[E] &= \Pr\big[E \;\big|\; D_i = 1, D_j = 0\big] \cdot \Pr[D_i = 1, D_j = 0] \\ &= \Pr\big[E' \;\big|\; D'_i = 1, D'_j = 0\big] \cdot \Pr[D'_i = 1, D'_j = 0] \\ &= \Pr[E'] .\end{aligned}$$

This finishes the proof of Lemma 4.1. □

Now we present the proof of Theorem 2.12. Consider a 2-local $(1,n)$-sampler S. If condition (1) from Lemma 4.1 does not hold for $S(U)$, then we can reject outright. Otherwise, there are five cases corresponding to the marginal probability that any particular coordinate is 1.

Case 0: If all output bits of S are constant 0 then $S(U)$ is trivially exchangeable.

Case 1/4: In this case, each edge of the graph is AND-type. When two AND-type edges share an endpoint, we say that they *agree* on the endpoint if the unique assignments that make the two edges evaluate to 1 agree on the value of the node. We assume without loss of generality that the graph has no nodes of degree 0. We claim that $S(U)$ is exchangeable iff at least one of the following conditions holds.

(i) The edges are all disjoint.

(ii) The graph is a star, and all edges agree on the central node.[4]

(iii) The graph is a triangle, and there is agreement at all nodes.[5]

(iv) The graph is a triangle, and there is disagreement at all nodes.

(v) There are only three nodes u, v, w, and there are no $\{u, w\}$ edges, there are at most two $\{u, v\}$ edges and they agree on v and disagree on u, there are at most two $\{v, w\}$ edges and they agree on v and disagree on w, and the $\{u, v\}$ edges disagree with the $\{v, w\}$ edges on v.

(vi) There are only two nodes, and no two edges agree on both nodes.

(vii) There are only two nodes, and all edges agree on both nodes.

It is trivial to check in linear time whether at least one of these conditions holds.

First we assume at least one of the conditions holds, and argue that $S(U)$ is exchangeable. If (i) holds then $S(U)$ is i. i. d. where each coordinate has $1/4$ probability of being 1, and this is exchangeable. If (ii) holds then $S(U)$ is a uniform mixture of the uniform distribution and the constant all 0's distribution, so $S(U)$ is exchangeable since it is a mixture of i. i. d.'s. If (iii) holds then outputs of Hamming weight 1 each

[4] A star does not include multi-edges.

[5] A triangle does not include multi-edges.

have probability $1/8$, and outputs of Hamming weight 2 each have probability 0, so $S(U)$ is exchangeable since the probability of an output only depends on the Hamming weight. If (iv) or (v) or (vi) holds then an edge evaluating to 1 forces all other edges to evaluate to 0, so $S(U)$ is all 0's with probability $1-(n/4)$ and is otherwise uniformly distributed on strings of Hamming weight 1, so $S(U)$ is exchangeable. If (vii) holds then $S(U)$ is all 1's with probability $1/4$ and all 0's with probability $3/4$, which is exchangeable.

We prove the converse with two lemmas, which show that the following conditions are the only obstacles to exchangeability. We write $\exists e_1, e_2, e_3$ with the tacit assumption that these are three *distinct* edges.

(A) $\exists e_1, e_2, e_3$ such that e_1, e_3 are not disjoint, and e_2, e_3 are disjoint.

(B) $\exists e_1, e_2, e_3$ all sharing an endpoint on which e_1, e_2 disagree, and such that e_3 does not share its other endpoint with e_1 or with e_2.

(C) $\exists e_1, e_2, e_3$ such that e_1, e_3 share both endpoints, and e_2 shares exactly one endpoint with them, and they all agree on the common node.

(D) $\exists e_1, e_2, e_3$ forming a triangle, such that e_1, e_3 agree and e_2, e_3 disagree.

(E) $\exists e_1, e_2, e_3$ such that e_1, e_3 share and agree on both endpoints, and e_2 either does not share both endpoints or does not agree on both endpoints.

Lemma 4.2. *If none of (i)–(vii) hold then at least one of (A)–(E) holds.*

Proof. Assume none of (i)–(vii) hold. Suppose the graph is not connected. Then since (i) fails, there are two edges that are not disjoint, and there is also another edge not in their connected component, so (A) holds. Henceforth suppose the graph is connected.

Suppose there are at least four nodes. If there is a simple path of length three, then (A) holds, so suppose there is no such path. Then the graph is "star-like," meaning it would be a star if multi-edges were replaced with single edges. If there is not complete agreement on the central node then (B) holds; otherwise, since (ii) fails, there must be multi-edges and so (C) holds.

Now suppose there are exactly three nodes and there is a triangle. If there are no multi-edges, then since (iii) and (iv) fail, (D) holds. If there is a multi-edge pair, then either these two edges disagree on some endpoint, in which case (D) holds, or they agree on both endpoints, in which case (E) holds.

Now suppose there are exactly three nodes and there is no triangle. Since the graph is connected, there is a length-2 path, say $\{u,v\},\{v,w\}$. Since (v) fails, either there are two $\{u,v\}$ edges that disagree on v, in which case (B) holds, or there are two $\{u,v\}$ edges that agree on both u and v, in which case (E) holds, or analogous situations happen with $\{v,w\}$, or all edges agree on v and there are either two $\{u,v\}$ edges or two $\{v,w\}$ edges, in which case (C) holds. Note that there cannot be just one $\{u,v\}$ edge and just one $\{v,w\}$ edge, since if they agreed then (ii) would hold, and if they disagreed then (v) would hold.

Finally, if there are exactly two nodes then since (vi) and (vii) fail, (E) holds. $\square$

Lemma 4.3. *If at least one of (A)–(E) holds, then $S(U)$ is not exchangeable.*

Proof. Assuming at least one of (A)–(E) holds, we use condition (2) from Lemma 4.1 to refute exchangeability of $S(U)$ by showing that the marginal probability that $e_3 = 1$ (more precisely, the random variable indexed by e_3 evaluates to 1) changes when we go from conditioning on $(e_1 = 1,\ e_2 = 0)$ to conditioning on $(e_1 = 0,\ e_2 = 1)$.

If (A) holds and e_1, e_3 share both endpoints then it goes either from 1 to 0 (if e_1, e_3 agree on both endpoints) or from 0 to $1/3$. If (A) holds and e_1, e_3 share only one endpoint and e_1, e_2 are disjoint then it goes either from $1/2$ to $1/6$ (if e_1, e_3 agree) or from 0 to $1/3$. If (A) holds and e_3, e_1, e_2 form a simple path then it goes either from $1/2$ to 0 (if e_1, e_3 agree and e_1, e_2 agree) or from $1/2$ to $1/4$ (if e_1, e_3 agree and e_1, e_2 disagree) or from 0 to $1/2$ (if e_1, e_3 disagree and e_1, e_2 agree) or from 0 to $1/4$.

If (B) holds then it goes either from $1/2$ to 0 (if e_1, e_3 agree) or from 0 to $1/2$. If (C) holds then it goes either from 1 to 0 (if e_1, e_3 agree on both endpoints) or from 0 to 1. If (D) holds then it goes either from 1 to 0 (if e_1, e_2 agree) or from $1/2$ to 0. If (E) holds then it goes from 1 to 0. $\square$

Interestingly, the above analysis shows that in Case 1/4, $S(U)$ is "globally exchangeable" iff it is "locally exchangeable" in the sense that every set of three coordinates is exchangeable.

Case 1/2: In this case, each edge of the graph is either XOR-type or a self-loop. We say that XOR-type multi-edges *agree* if they compute the same function, and similarly we say that multi-self-loops agree if they compute the same function. We assume without loss of generality that the graph has no nodes of degree 0. We claim that $S(U)$ is exchangeable iff at least one of the following conditions holds.

(i) The graph is a forest ignoring self-loops, and there is at most one self-loop per connected component.

(ii) The graph is a simple cycle.[6]

(iii) The graph is a simple path but with two self-loops, one at each end.

(iv) There are only two nodes, no self-loops, and all edges agree.

(v) There are only two nodes, no self-loops, and only two edges, which disagree.

(vi) There is only one node, with agreeing self-loops.

(vii) There is only one node, with only two self-loops, which disagree.

It is trivial to check in linear time whether at least one of these conditions holds.

First we assume at least one of the conditions holds, and argue that $S(U)$ is exchangeable. If (i) holds then $S(U)$ is uniform by the characterization in the proof of Theorem 2.11. If (ii) or (iii) holds then $S(U)$ is the same as conditioning the uniform distribution on having a particular parity, so $S(U)$ is invariant under permuting coordinates (by commutativity and associativity of addition over $GF(2)$). If (iv) or (vi) holds then $S(U)$ is all 1's with probability $1/2$ and all 0's with probability $1/2$, which is exchangeable. If (v) or (vii) holds then $S(U)$ is uniform over the two possibilities 01 and 10, which is exchangeable.

We prove the converse with two lemmas, which show that the following conditions are the only obstacles to exchangeability.

[6] A simple cycle does not include multi-edges or self-loops, unless it has length 2, in which case it is a multi-edge pair.

(A) $\exists$ a cycle C of XOR-type edges, and an edge e that is either a self-loop or has at least one endpoint not on C.

(B) $\exists$ a path P (of zero or more XOR-type edges), two self-loops with one at each end of P, and an edge e at least one of whose nodes is not on P.

(C) $\exists$ three XOR-type edges all sharing both endpoints, such that some but not all of these edges agree.

(D) $\exists$ three self-loops all sharing the same node, such that some but not all of these edges agree.

Lemma 4.4. *If none of (i)–(vii) hold then at least one of (A)–(D) holds.*

Proof. Assume none of (i)–(vii) hold. If there is only one node, then since (vi) and (vii) fail, (D) holds. Henceforth suppose there are at least two nodes. Since (i) fails, the graph has either a cycle of XOR-type edges, or two self-loops in the same connected component.

If it has a cycle of XOR-type edges, then let C be a shortest such cycle. Since (ii) fails, there exists another edge. If there exists another edge e that is either a self-loop or has at least one endpoint not on C, then (A) holds. Otherwise, all other edges are XOR-type with both endpoints on C. If C had length at least three, this would contradict the minimal nature of C. Hence there are only two nodes, with no self-loops and with at least three XOR-type edges (two of them forming C). Then since (iv) and (v) fail, (C) holds.

On the other hand, if the graph is a forest ignoring self-loops but has two self-loops in the same connected component, then consider two such self-loops that are closest, and let P be the unique path between them. If P has length zero (so the self-loops are at the same node), then (B) holds since we are assuming there are at least two nodes (and the other node is incident to some edge e). Otherwise, since (iii) fails, there exists another edge e. If e were XOR-type with both endpoints on P, then there would be a cycle of XOR-type edges (contradicting the assumption that the graph is a forest ignoring self-loops), and if e were another self-loop on P then this would contradict the minimal nature of P (since P has length ≥ 1). Thus at least one of e's nodes is not on P, so (B) holds. $\square$

Lemma 4.5. *If at least one of (A)–(D) holds, then $S(U)$ is not exchangeable.*

Proof. Assuming at least one of (A)–(D) holds, we use condition (2) from Lemma 4.1 to refute exchangeability of $S(U)$ by exhibiting edges e_1, e_2 for which the joint distribution of the evaluations of the other edges changes when we go from conditioning on $(e_1 = 1,\ e_2 = 0)$ to conditioning on $(e_1 = 0,\ e_2 = 1)$.

If (A) holds then let $e_1 = e$ and e_2 be any edge on C. The joint distribution of the other edges on C (besides e_2) goes from being uniform conditioned on having a particular parity to being uniform conditioned on having the opposite parity.

If (B) holds then let $e_1 = e$ and e_2 be one of the two self-loops at the ends of P. The joint distribution of the other edges on P, together with the self-loop at the other end of P, goes from being uniform conditioned on having a particular parity to being uniform conditioned on having the opposite parity.

If (C) or (D) holds then call the edges e_1, e_2, e_3 where e_1, e_2 disagree and e_2, e_3 agree. Then the marginal probability that e_3 evaluates to 1 goes from 0 to 1. $\square$

Case 3/4: In this case, each edge of the graph is OR-type. Let $\overline{S}$ denote the circuit obtained from S by negating every output bit. Then $S(U)$ is exchangeable iff $\overline{S}(U)$ is exchangeable. Every edge of the graph for $\overline{S}$ is AND-type, so we can use the characterization from Case 1/4 to decide in linear time whether $S(U)$ is exchangeable.

Case 1: If all output bits of S are constant 1 then $S(U)$ is trivially exchangeable.

4.3 Pairwise independence for 2-local samplers

We prove Theorem 2.13. Consider a 2-local $(1,n)$-sampler S, and assume without loss of generality that S has no constant output bits. We claim that $S(U)$ is pairwise independent iff both of the following conditions hold.

(i) The graph has no multi-edges.

(ii) For each node v of the graph, there is at most one of the following among the edges incident to v: a self-loop, an AND-type edge, or an OR-type edge.

It is trivial to check in linear time whether condition (ii) holds. Condition (i) is an instance of the element distinctness problem, which is the problem of deciding whether a list of numbers (encoding pairs of nodes, in our situation) has no duplicates. The element distinctness problem can be solved in deterministic $O(n \log n)$ time by sorting, and it can be solved in zero-error randomized expected linear time.[7] We supply a folklore proof of the following lemma in Section A.2.

Lemma 4.6. *The element distinctness problem has a zero-error randomized expected linear-time algorithm.*

We now verify that (i) and (ii) characterize pairwise independence. First we assume that (i) and (ii) both hold, and show that the evaluations of two arbitrary edges e_1, e_2 are independent. If e_1, e_2 are disjoint then this is immediate; otherwise they share a node v. Since (i) and (ii) hold, the characterization in the proof of Theorem 2.11 implies that the edges incident to v are fully independent of each other; in particular e_1, e_2 are independent. Conversely, suppose (i) and (ii) do not both hold. A simple case analysis shows that if two edges form a multi-edge pair, or if they share a node and neither is XOR-type, then they cannot be independent, and so $S(U)$ is not pairwise independent.

5 Proofs of approximate completeness results

We prove Theorem 2.16 and Theorem 2.17 in Section 5.1, and we prove Theorem 2.19 and Theorem 2.20 in Section 5.2.

[7] As usual, we assume a model of computation where arithmetic operations and array look-ups take constant time.

Definition 5.1. For functions $0 \le c(s) < f(s) \le 1$, STATISTICAL-DISTANCE^{c,f} is the following promise problem.

$$\begin{aligned}
\text{STATISTICAL-DISTANCE}^{c,f}_{\text{YES}} &= \big\{S \,:\, S \text{ is a } (k,2,s)\text{-sampler and } S(U)_1, S(U)_2 \text{ are } c(s)\text{-close}\big\}, \\
\text{STATISTICAL-DISTANCE}^{c,f}_{\text{NO}} &= \big\{S \,:\, S \text{ is a } (k,2,s)\text{-sampler and } S(U)_1, S(U)_2 \text{ are } f(s)\text{-far}\big\}.
\end{aligned}$$

Without loss of generality, $S(U)$ is independent.

Sahai and Vadhan [31] proved the following two theorems.

Theorem 5.2. STATISTICAL-DISTANCE^{c,f} *is* prSZK*-hard for all constants* $0 < c < f < 1$.

Theorem 5.3. STATISTICAL-DISTANCE$^{c,f} \in$ prSZK *for all constants* $0 < c < f^2 < 1$.

More generally, Sahai and Vadhan proved that for all functions c, f computable in time polynomial in s, the problem STATISTICAL-DISTANCE^{c,f} is prSZK-hard if $c = 2^{-s^{o(1)}}$, $f = 1 - 2^{-s^{o(1)}}$, and is in prSZK if $c \le f^2 - s^{-O(1)}$. This can be used to improve the parameters (as functions of s) in our theorems. For example, in Theorem 2.16, c can be $2^{-s^{o(1)}}$ and f can be $(1/4) - 2^{-s^{o(1)}}$. We chose to state our theorems using constants (except Theorem 2.17, where the $1/(n+1)$ factor is needed) for simplicity and clarity. It is awkward to handle reductions when the c, f functions depend on the size of one circuit for one problem but on the size of a different circuit for the other problem.

5.1 Approximate full independence

We now prove Theorem 2.16 and Theorem 2.17.

Proof of Theorem 2.16. We reduce STATISTICAL-DISTANCE^{2c,4f} (which is prSZK-hard according to Theorem 5.2) to FULLY-INDEPENDENT^{c,f}. Given a $(k,2)$-sampler S, let $D = S(U)$. The reduction outputs a sampler S' that outputs (b, w) where $b \in \{1,2\}$ is chosen uniformly at random and $w \in \{0,1\}^k$ is a sample from D_b. This yields a distribution $D' = S'(U)$ over $\{1,2\} \times \{0,1\}^k$, but D' can be viewed as a distribution over $(\{0,1\}^k)^2$ by embedding $\{1,2\}$ into $\{0,1\}^k$.

First we show that if $S \in$ STATISTICAL-DISTANCE$^{2c,4f}_{\text{YES}}$ then $S' \in$ FULLY-INDEPENDENT$^{c,f}_{\text{YES}}$. Suppose $\|D_1 - D_2\| \le 2c$. Let D^* be the independent distribution whose first coordinate is uniform over $\{1,2\}$ and whose second coordinate is D_1. For any event $E \subseteq \{1,2\} \times \{0,1\}^k$ and $b \in \{1,2\}$, let $E_b = \big\{z \in \{0,1\}^k \,:\, (b,z) \in E\big\}$. We have

$$\Pr[D' \in E] = \frac{1}{2} \cdot \Pr[D_1 \in E_1] + \frac{1}{2} \cdot \Pr[D_2 \in E_2] \quad \text{and} \quad \Pr[D^* \in E] = \frac{1}{2} \cdot \Pr[D_1 \in E_1] + \frac{1}{2} \cdot \Pr[D_1 \in E_2].$$

This implies that

$$\big|\Pr[D' \in E] - \Pr[D^* \in E]\big| \;=\; \frac{1}{2} \cdot \big|\Pr[D_2 \in E_2] - \Pr[D_1 \in E_2]\big| \;\le\; \frac{1}{2} \cdot 2c \;=\; c$$

where the inequality follows by $\|D_1 - D_2\| \le 2c$. Hence $\|D' - D^*\| \le c$.

Now we show that if $S \in \text{STATISTICAL-DISTANCE}^{2c,4f}_{\text{NO}}$ then $S' \in \text{FULLY-INDEPENDENT}^{c,f}_{\text{NO}}$. Suppose $\|D' - D^*\| < f$ for some distribution D^* (not necessarily the same as above) that is independent. Assume $\Pr[D^*_1 = 1] \leq \Pr[D^*_1 = 2]$ (the other case is symmetric). For any event $E \subseteq \{0,1\}^k$, we have

$$\begin{aligned}
\Pr[D_1 \in E] &= 2 \cdot \Pr\big[D' \in \{1\} \times E\big] \\
&< 2 \cdot \big(\Pr\big[D^* \in \{1\} \times E\big] + f\big) \\
&\leq 2 \cdot \big(\Pr\big[D^* \in \{2\} \times E\big] + f\big) \\
&< 2 \cdot \big(\Pr\big[D' \in \{2\} \times E\big] + 2f\big) \\
&= 2 \cdot \big((1/2) \cdot \Pr[D_2 \in E] + 2f\big) \\
&= \Pr[D_2 \in E] + 4f
\end{aligned}$$

where the third line follows by independence and by the assumption that $\Pr[D^*_1 = 1] \leq \Pr[D^*_1 = 2]$. Hence by the second equality in Definition 2.14, $\|D_1 - D_2\| < 4f$. □

The proof of Theorem 2.17 uses the following lemma.

Lemma 5.4. *Suppose D is a distribution over $(\{0,1\}^k)^n$. If D is c-close to some fully independent distribution D^*, then D is $(n+1)c$-close to the distribution D' that is fully independent and has the same marginals as D.*

Lemma 5.4 can be proven using a simple hybrid argument. The case $n = 2$ was proven in [5], but the same argument works for general n; we omit the details.

Proof of Theorem 2.17. We reduce $\text{FULLY-INDEPENDENT}^{c,f}$ to $\text{STATISTICAL-DISTANCE}^{c',f}$ (which is in prSZK by Theorem 5.3). Given a (k,n)-sampler S, construct a (k,n)-sampler S' that runs S independently n times and outputs $(w_1, \ldots, w_n)$ where w_i is the i^{th} coordinate of the output of the i^{th} run. Let $D = S(U)$ and $D' = S'(U)$, and note that D' is as in the statement of Lemma 5.4.

The reduction outputs a $(kn,2)$-sampler $\hat{S}$ whose first coordinate is a sample from D and whose second coordinate is a sample from D'. If $S \in \text{FULLY-INDEPENDENT}^{c,f}_{\text{YES}}$ then by Lemma 5.4,

$$\|D - D'\| \leq (n+1)c = c'$$

and hence $\hat{S} \in \text{STATISTICAL-DISTANCE}^{c',f}_{\text{YES}}$. If $S \in \text{FULLY-INDEPENDENT}^{c,f}_{\text{NO}}$ then $\|D - D'\| \geq f$ since D' is fully independent, and hence $\hat{S} \in \text{STATISTICAL-DISTANCE}^{c',f}_{\text{NO}}$. □

5.2 Approximate exchangeability

We now prove Theorem 2.19 and Theorem 2.20.

Proof of Theorem 2.19. We reduce $\text{STATISTICAL-DISTANCE}^{c,2f}$ (which is prSZK-hard by Theorem 5.2) to $\text{EXCHANGEABLE}^{c,f}$. Given a $(k,2)$-sampler S and letting $D = S(U)$ where, without loss of generality, D_1, D_2 are independent, the reduction is the identity map.

First we show that if $S \in \text{STATISTICAL-DISTANCE}^{c,2f}_{\text{YES}}$ then $S \in \text{EXCHANGEABLE}^{c,f}_{\text{YES}}$. Suppose $\|D_1 - D_2\| \leq c$. Let D^* be the independent distribution over $(\{0,1\}^k)^2$ both of whose marginals are

D_1. Since D^* is i. i. d., it is exchangeable. For any event $E \subseteq (\{0,1\}^k)^2$ and $y \in \{0,1\}^k$, let $E_y = \{z \in \{0,1\}^k : (y,z) \in E\}$. By independence, we have

$$\Pr[D \in E] = \sum_{y \in \{0,1\}^k} \Pr[D_1 = y] \cdot \Pr[D_2 \in E_y]$$

and

$$\Pr[D^* \in E] = \sum_{y \in \{0,1\}^k} \Pr[D_1 = y] \cdot \Pr[D_1 \in E_y] \, .$$

This implies that

$$\begin{aligned} \big|\Pr[D \in E] - \Pr[D^* \in E]\big| &\leq \sum_{y \in \{0,1\}^k} \Pr[D_1 = y] \cdot \big|\Pr[D_2 \in E_y] - \Pr[D_1 \in E_y]\big| \\ &\leq \sum_{y \in \{0,1\}^k} \Pr[D_1 = y] \cdot c \\ &= c \end{aligned}$$

where the second inequality follows by $\|D_1 - D_2\| \leq c$. Hence $\|D - D^*\| \leq c$.

Now we show that if $S \in \text{STATISTICAL-DISTANCE}_{\text{NO}}^{c,2f}$ then $S \in \text{EXCHANGEABLE}_{\text{NO}}^{c,f}$. Suppose $\|D - D^*\| < f$ for some distribution D^* (not necessarily the same as above) that is exchangeable. In particular, D_1^*, D_2^* are identically distributed. We trivially have $\|D_1 - D_1^*\| \leq \|D - D^*\|$ and $\|D_2 - D_2^*\| \leq \|D - D^*\|$. Thus by the triangle inequality, $\|D_1 - D_2\| \leq \|D_1 - D_1^*\| + \|D_2 - D_2^*\| < 2f$. □

The proof of Theorem 2.20 uses the following lemma.

Lemma 5.5. *Suppose D is a distribution over $(\{0,1\}^k)^n$. If D is c-close to some exchangeable distribution D^*, then D is $2c$-close to the distribution D' obtained by drawing a sample from D then permuting the coordinates according to a uniformly random permutation.*

Proof of Lemma 5.5. For a multiset $W \subseteq \{0,1\}^k$ of size n, we say that $w \in (\{0,1\}^k)^n$ is an *ordering* of W if the multiset $\{w_i : i \in \{1,\dots,n\}\}$ equals W. Let $\mathrm{Ord}(W)$ denote the set of all orderings of W. Let d_W^{*+} be the sum of $\Pr[D = w] - \Pr[D^* = w]$ over all $w \in \mathrm{Ord}(W)$ such that $\Pr[D = w] - \Pr[D^* = w] > 0$, and let d_W^{*-} be the sum of $\Pr[D^* = w] - \Pr[D = w]$ over all $w \in \mathrm{Ord}(W)$ such that $\Pr[D^* = w] - \Pr[D = w] > 0$. Then by the third equality in Definition 2.14, we have

$$\|D - D^*\| = \sum_{\substack{\text{multisets } W \subseteq \{0,1\}^k \\ \text{of size } n}} \frac{d_W^{*+} + d_W^{*-}}{2} \, . \tag{5.1}$$

Letting $d_W'^{+}$ and $d_W'^{-}$ be the analogous quantities with D' instead of D^*, we have

$$\|D - D'\| = \sum_{\substack{\text{multisets } W \subseteq \{0,1\}^k \\ \text{of size } n}} \frac{d_W'^{+} + d_W'^{-}}{2} \, . \tag{5.2}$$

Now fix some W. Note that since D^* is exchangeable, all elements of $\mathrm{Ord}(W)$ have the same probability under D^*; call this probability p^*_W. If w is an element of $\mathrm{Ord}(W)$ then permuting the coordinates of w uniformly at random yields a uniformly random element of $\mathrm{Ord}(W)$. Thus all elements of $\mathrm{Ord}(W)$ have the same probability under D', namely

$$p'_W \;=\; \frac{1}{|\mathrm{Ord}(W)|}\cdot\sum_{w\in\mathrm{Ord}(W)}\Pr[D=w].$$

If $p'_W \geq p^*_W$ then $d'^{+}_W \leq d^{*+}_W$ by definition. If $p'_W \leq p^*_W$ then $d'^{-}_W \leq d^{*-}_W$ by definition. We also have

$$\begin{aligned} 0 \;&=\; \Big(\sum_{w\in\mathrm{Ord}(W)}\Pr[D=w]\Big) - |\mathrm{Ord}(W)|\cdot p'_W \\ &=\; \sum_{w\in\mathrm{Ord}(W)}\big(\Pr[D=w]-p'_W\big) \\ &=\; d'^{+}_W - d'^{-}_W \end{aligned}$$

which implies that $d'^{+}_W = d'^{-}_W \leq \max(d^{*+}_W, d^{*-}_W)$. Hence $\big(d'^{+}_W + d'^{-}_W\big) \leq 2\cdot\max(d^{*+}_W, d^{*-}_W) \leq 2\cdot\big(d^{*+}_W + d^{*-}_W\big)$. Since this holds for all W, we get $\|D-D'\| \leq 2\cdot\|D-D^*\|$ by (5.1) and (5.2). $\square$

We mention that the constant factor of 2 in Lemma 5.5 is tight, by the following example. Suppose $k=1$, and suppose D is uniformly distributed over a set of n strings, one of which has Hamming weight 1 and the other $n-1$ of which have Hamming weight $n-1$. Let D^* be uniformly distributed over the strings of Hamming weight $n-1$. Note that D^* is exchangeable, and $\|D-D^*\| = 1/n$. However, D' has probability $1/n^2$ on each string of Hamming weight 1, and probability $(n-1)/n^2$ on each string of Hamming weight $n-1$, and thus

$$\|D-D'\| = 2\Big(1-\frac{1}{n}\Big)\cdot\frac{1}{n} = 2\Big(1-\frac{1}{n}\Big)\cdot\|D-D^*\|\,.$$

Proof of Theorem 2.20. For any constant c' such that $2c < c' < f^2$, we reduce EXCHANGEABLE^{c,f} to STATISTICAL-DISTANCE$^{c',f}$ (which is in prSZK by Theorem 5.3). Given a (k,n)-sampler S, construct a (k,n)-sampler S'' that performs the following computation.

```
for i = 1, ..., ⌈log_2(1/(c'−2c))⌉ do
    choose π ∈ {0,1}^⌈log_2(n!)⌉ uniformly at random
    if π < n! then
        interpret π as a permutation on {1, ..., n}
        run S to get w = (w_1, ..., w_n)
        halt and output (w_π(1), ..., w_π(n))
    end
end
halt and output the all 0's element of ({0,1}^k)^n
```

Let $D = S(U)$, let D' be as in the statement of Lemma 5.5, and let $D'' = S''(U)$. Conditioned on halting inside the **for** loop, D'' has the same distribution as D'. In each iteration, there is $> 1/2$ probability that $\pi < n!$ and the computation of S'' halts. Hence the probability the computation halts on the last line (after failing to halt inside the **for** loop) is $< c' - 2c$. This implies that $\|D' - D''\| < c' - 2c$.

The reduction outputs a $(kn, 2)$-sampler $\hat{S}$ whose first coordinate is a sample from D and whose second coordinate is a sample from D''. If $S \in \text{EXCHANGEABLE}^{c,f}_{\text{YES}}$ then by Lemma 5.5, $\|D - D'\| \le 2c$, so by the triangle inequality $\|D - D''\| \le \|D - D'\| + \|D' - D''\| < 2c + (c' - 2c) = c'$ and hence $\hat{S} \in \text{STATISTICAL-DISTANCE}^{c',f}_{\text{YES}}$. If $S \in \text{EXCHANGEABLE}^{c,f}_{\text{NO}}$ then $\|D - D''\| \ge f$ since D'' is exchangeable, and hence $\hat{S} \in \text{STATISTICAL-DISTANCE}^{c',f}_{\text{NO}}$. □

6 $\mathsf{BC_{=}P}$

We now consider the bounded-error version of $\mathsf{C_{=}P}$, which does not seem to have been defined or studied in the literature before.[8]

Definition 6.1. $\mathsf{prBC_{=}P}$ is the class of all promise problems Karp-reducible to the following promise problem, BOUNDED-UNIFORM-BIT.

$$\begin{aligned} \text{BOUNDED-UNIFORM-BIT}_{\text{YES}} &= \{S : S \text{ is a } (1,1)\text{-sampler and } S(U) \text{ is uniform}\}, \\ \text{BOUNDED-UNIFORM-BIT}_{\text{NO}} &= \{S : S \text{ is a } (1,1)\text{-sampler and } \left|\Pr[S(U) = 1] - (1/2)\right| \ge 1/4\}. \end{aligned}$$

$\mathsf{BC_{=}P}$ is defined as the class of languages in $\mathsf{prBC_{=}P}$.

We now investigate the structural properties of $\mathsf{BC_{=}P}$. We begin with the following amplification result for $\mathsf{BC_{=}P}$, which is somewhat less trivial than usual amplification results. Then we apply this lemma to obtain closure properties of $\mathsf{BC_{=}P}$.

Lemma 6.2. *For all languages L, the following are equivalent.*

(1) For some polynomial q, there is a Karp reduction that takes x and outputs a $(1,1)$-sampler S such that the following both hold.[9]

$$\begin{aligned} x \in L &\implies S(U) \text{ is uniform}. \\ x \notin L &\implies \left|\Pr[S(U) = 1] - \frac{1}{2}\right| \ge \frac{1}{q(|x|)}. \end{aligned}$$

(2) $L \in \mathsf{BC_{=}P}$.

[8] A more extreme version, in which YES instances have acceptance probability 1/2 and NO instances have acceptance probability 0, has been studied before and is known as $\mathsf{C_{==}P[half]}$ [9] and HalfP [12].

[9] It is not convenient to phrase this as a reduction *to a problem*, because the bound $\frac{1}{q(|x|)}$ depends on the size of x, not the size of S.

(3) For every polynomial Q, there is a Karp reduction that takes x and outputs a $(1,1)$-sampler S such that the following both hold.

$$x \in L \implies S(U) \text{ is uniform}.$$
$$x \notin L \implies \Pr[S(U)=1] \leq \frac{1}{2^{Q(|x|)}}.$$

Proof. Clearly $(3) \Rightarrow (2) \Rightarrow (1)$, so we just need to demonstrate $(1) \Rightarrow (3)$. Assume (1). By the standard trick for making $\mathsf{C_{=}P}$ have "1-sided error" (see Section A.1), it follows that there is a similar reduction mapping x to some (different) $(1,1)$-sampler S that achieves

$$\Pr[S(U)=1] \leq \frac{1}{2} - \frac{1}{q(|x|)^2}$$

in the case $x \notin L$. Construct a new circuit S' that performs the following computation.

> let $m = 2q(|x|)^4 Q(|x|)$ and $t = \left(\frac{1}{2} - \frac{1}{2q(|x|)^2}\right) \cdot m$
> choose a uniformly random bit b
> **if** $b = 0$ **then** run S independently m times, and accept iff $\geq t$ of these runs accept **else** accept with probability $1 - \frac{1}{2^m}\sum_{i=\lceil t\rceil}^{m}\binom{m}{i}$

Note that the values of m, t, and $\sum_{i=\lceil t\rceil}^{m}\binom{m}{i}$ can be precomputed in polynomial time and hard-wired into S'. If S accepts with probability $1/2$ then the probability that $\geq t$ of m runs of S accept is

$$\frac{1}{2^m}\sum_{i=\lceil t\rceil}^{m}\binom{m}{i}.$$

Hence if $x \in L$ then S' accepts with probability $1/2$. If S accepts with probability

$$\leq \frac{1}{2} - \frac{1}{q(|x|)^2}$$

then by a standard concentration bound, the probability that $\geq t$ of m runs of S accept is

$$\leq 2^{-m/2q(|x|)^4} = \frac{1}{2^{Q(|x|)}}.$$

Also by a standard concentration bound,

$$\left(1 - \frac{1}{2^m}\sum_{i=\lceil t\rceil}^{m}\binom{m}{i}\right) \leq 2^{-m/2q(|x|)^4} = \frac{1}{2^{Q(|x|)}}.$$

Hence if $x \notin L$ then S' accepts with probability

$$\leq \frac{1}{2}\cdot\frac{1}{2^{Q(|x|)}} + \frac{1}{2}\cdot\frac{1}{2^{Q(|x|)}} = \frac{1}{2^{Q(|x|)}}.$$

The reduction for (3) just outputs S'. $\square$

Note that coRP $\subseteq$ $\mathsf{BC_{=}P}$. The 1-sided error property in part (3) of Lemma 6.2 implies that $\mathsf{BC_{=}P} \subseteq$ BPP. Thus $\mathsf{BC_{=}P}$ is presumably closed under complement (since presumably $\mathsf{P} = \mathsf{BC_{=}P} = \mathsf{BPP}$ [23]), but proving this seems out of reach. We now apply Lemma 6.2 to prove that $\mathsf{BC_{=}P}$ is closed under union, intersection, disjunction (i. e., $\vee L \in \mathsf{BC_{=}P}$ if $L \in \mathsf{BC_{=}P}$, where $\vee L = \{(x_1,\ldots,x_\ell) : x_i \in L \text{ for some } i\}$), and conjunction (i. e., $\wedge L \in \mathsf{BC_{=}P}$ if $L \in \mathsf{BC_{=}P}$, where $\wedge L = \{(x_1,\ldots,x_\ell) : x_i \in L \text{ for all } i\}$). The proof of Lemma 3.4 in Section A.1 showing that $\mathsf{C_{=}P}$ is closed under disjunction does not work to show that $\mathsf{BC_{=}P}$ is closed under disjunction.

Theorem 6.3. $\mathsf{BC_{=}P}$ *is closed under disjunction.*

Proof. Assuming $L \in \mathsf{BC_{=}P}$, we exhibit a reduction witnessing $\vee L \in \mathsf{BC_{=}P}$. Given $(x_1,\ldots,x_\ell)$, by padding we may assume without loss of generality that the x_i's all have the same length $n \geq \ell$. By $(2) \Rightarrow (3)$ in Lemma 6.2, there is a reduction that takes x_i and outputs C_i such that the following both hold.

$$\begin{aligned} x_i \in L &\implies \Pr[C_i(U)=1] = \frac{1}{2}, \\ x_i \notin L &\implies \Pr[C_i(U)=1] \leq \frac{1}{4n}. \end{aligned}$$

Our reduction witnessing $\vee L \in \mathsf{BC_{=}P}$ runs the above reduction for L on each x_i to obtain the circuits C_i, then outputs a circuit S that runs each C_i independently and combines their results with a parity gate. If $(x_1,\ldots,x_\ell) \in \vee L$ then S is taking the parity of ℓ independent bits at least one of which is uniform, so $S(U)$ is uniform. If $(x_1,\ldots,x_\ell) \notin \vee L$ then letting $z_1,\ldots,z_\ell$ denote S's random input, we have

$$\begin{aligned} \Pr[S(U)=1] &= \Pr_{z_1,\ldots,z_\ell}\Big[\bigoplus_{i=1}^{\ell} C_i(z_i) = 1\Big] \\ &\leq \Pr_{z_1,\ldots,z_\ell}\big[C_i(z_i) = 1 \text{ for some } i\big] \\ &\leq \sum_{i=1}^{\ell} \Pr_{z_i}[C_i(z_i)=1] \\ &\leq \ell \cdot \frac{1}{4n} \\ &\leq \frac{1}{4}. \end{aligned}$$

□

Theorem 6.4. $\mathsf{BC_{=}P}$ *is closed under conjunction.*

Proof. Assuming $L \in \mathsf{BC_{=}P}$, we exhibit a reduction witnessing $\wedge L \in \mathsf{BC_{=}P}$. We are given $(x_1,\ldots,x_\ell)$. Lemma 6.2 implies that $\mathsf{BC_{=}P}$ can have 1-sided error, so there is a reduction that takes x_i and outputs C_i such that the following both hold.

$$\begin{aligned} x_i \in L &\implies \Pr[C_i(U)=1] = \frac{1}{2}, \\ x_i \notin L &\implies \Pr[C_i(U)=1] \leq \frac{1}{4}. \end{aligned}$$

Our reduction witnessing $\wedge L \in \mathsf{BC_=P}$ runs the above reduction for L on each x_i to obtain the circuits C_i, then outputs a circuit S that chooses $i \in \{1, \ldots, \ell\}$ uniformly at random and outputs the same bit as an execution of C_i. Thus

$$\Pr[S(U) = 1] = \frac{1}{\ell} \sum_{i=1}^{\ell} \Pr[C_i(U) = 1]\,.$$

If $(x_1, \ldots, x_\ell) \in \wedge L$ then $S(U)$ is uniform since $\Pr[C_i(U) = 1] = 1/2$ for all i. If $(x_1, \ldots, x_\ell) \notin \wedge L$ then

$$\Pr[S(U) = 1] \leq \frac{\ell - 1}{\ell} \cdot \frac{1}{2} + \frac{1}{\ell} \cdot \frac{1}{4} = \frac{1}{2} - \frac{1}{4\ell}\,.$$

By $(1) \Rightarrow (2)$ in Lemma 6.2, $\wedge L \in \mathsf{BC_=P}$. $\square$

Corollary 6.5. $\mathsf{BC_=P}$ *is closed under union and intersection.*

Proof. This follows by the same proof techniques used for Theorem 6.3 and Theorem 6.4. Alternatively, this follows in a black-box fashion from Theorem 6.3 and Theorem 6.4: Assuming $L_1 \in \mathsf{BC_=P}$ and $L_2 \in \mathsf{BC_=P}$, we also have $L \in \mathsf{BC_=P}$ where $L = \big\{(x,i) \;:\; i \in \{1,2\} \text{ and } x \in L_i\big\}$, and thus $\vee L \in \mathsf{BC_=P}$ and $\wedge L \in \mathsf{BC_=P}$. But $L_1 \cup L_2$ reduces to $\vee L$, and $L_1 \cap L_2$ reduces to $\wedge L$, both by the same trivial reduction that maps x to $\big((x,1),(x,2)\big)$. Hence $L_1 \cup L_2 \in \mathsf{BC_=P}$ and $L_1 \cap L_2 \in \mathsf{BC_=P}$. $\square$

7 Open problems

There are plenty of open problems concerning the complexity of deciding statistical properties of joint distributions with low-complexity samplers. To the best of our knowledge, none of these problems has ever been studied before, so the best bounds are what hold trivially or follow directly from results discussed in this work.

What can be said about depth-2 circuits? For example, for each of the languages listed in Fact 2.3, it is consistent with current knowledge that, when restricted to depth-2 circuit samplers, the language could be in P or $\mathsf{C_=P}$-complete. What about d-local samplers when $d > 2$ or $k > 1$? For example, is there a polynomial-time algorithm for deciding whether the joint distribution of a given 3-local $(1,n)$-sampler is fully independent? What are the complexities of deciding full independence and exchangeability for logspace samplers? What about samplers where each input bit influences at most a bounded number of output bits?

There are also plenty of open problems concerning the complexity of approximately deciding statistical properties of samplable distributions. Can quantitative improvements in our results be obtained (e. g., improvements in the bounds of $1/4$ in Theorem 2.16 and $1/2$ in Theorem 2.19)? What is the complexity of the exact-versus-far (as opposed to close-versus-far) versions of these problems? What about deciding whether a joint distribution is close or far from being pairwise independent? What can be said about versions of these problems where k is constrained to be 1? What about approximate problems restricted to low-complexity samplers? Can we prove any interesting approximate algorithmic results?

For general circuit samplers, what is the complexity of deciding whether or not a distribution is a mixture of i. i. d. distributions? It is not clear whether this problem is in $\mathsf{C_=P}$. Of course, i. i. d. corresponds to drawing balls from an urn with replacement, and when changed to *without* replacement the problem becomes equivalent to exchangeability, which is decidable in $\mathsf{C_{=}P}$.

Are there other interesting structural properties or applications of $\mathsf{BC_{=}P}$?

Acknowledgments

I thank László Babai and anonymous reviewers for their comments.

A Folklore proofs

We use this appendix to supply some folklore proofs.

A.1 Closure of $\mathsf{C_{=}P}$

We supply a folklore proof of Lemma 3.4, which is used in the proof of Theorem 2.10. Assume $L \in \mathsf{C_{=}P}$. First we prove that $\forall_q L \in \mathsf{C_{=}P}$.

There is a deterministic polynomial-time algorithm that takes as input x and outputs a circuit $C(y,z)$ where $|y| = q(|x|)$, such that the following both hold.

$$x \in \forall_q L \implies \forall y : \Pr_z[C(y,z) = 1] = \frac{1}{2},$$
$$x \notin \forall_q L \implies \exists y : \Pr_z[C(y,z) = 1] \neq \frac{1}{2}.$$

The circuit C runs the reduction witnessing $L \in \mathsf{C_{=}P}$ on (x,y), then runs the output of the reduction on z. Now let $\overline{C}$ be the same as C but with the output negated. We construct a new circuit $A(y,z')$ that for all y satisfies

$$\Pr_{z'}[A(y,z') = 1] \;=\; \frac{1}{2} \cdot \left(\Pr_z[C(y,z) = 1]^2 + \Pr_z[\overline{C}(y,z) = 1]^2\right)$$

by randomly choosing either C or $\overline{C}$ and running it twice independently. We also construct a new circuit $B(y,z')$ that for all y satisfies

$$\Pr_{z'}[B(y,z') = 1] \;=\; \frac{1}{2} \cdot \left(2 \cdot \Pr_z[C(y,z) = 1] \cdot \Pr_z[\overline{C}(y,z) = 1]\right)$$

by running both C and $\overline{C}$ independently. Thus we have

$$\Pr_{z'}[A(y,z') = 1] - \Pr_{z'}[B(y,z') = 1] \;=\; \frac{1}{2} \cdot \left(\Pr_z[C(y,z) = 1] - \Pr_z[\overline{C}(y,z) = 1]\right)^2$$

which implies the following.

$$x \in \forall_q L \implies \forall y: \ \Pr_{z'}[A(y,z')=1] - \Pr_{z'}[B(y,z')=1] = 0\,.$$

$$x \notin \forall_q L \implies \begin{cases} \exists y: \ \Pr_{z'}[A(y,z')=1] - \Pr_{z'}[B(y,z')=1] > 0 \\ \text{ and} \\ \forall y: \ \Pr_{z'}[A(y,z')=1] - \Pr_{z'}[B(y,z')=1] \geq 0\,. \end{cases}$$

Now we construct a circuit $S(y,z'')$ such that

$$\Pr_{z''}[S(y,z'')=1] \;=\; \frac{1}{2}\cdot\big(\Pr_{z'}[A(y,z')=1] + (1-\Pr_{z'}[B(y,z')=1])\big)$$

by randomly choosing either A or B to run, and negating the output if B was chosen. This implies the following.

$$x \in \forall_q L \implies \forall y: \ \Pr_{z''}[S(y,z'')=1] = \frac{1}{2} \qquad \implies \Pr_{y,z''}[S(y,z'')=1] = \frac{1}{2}\,.$$

$$x \notin \forall_q L \implies \begin{cases} \exists y: \ \Pr_{z''}[S(y,z'')=1] > \frac{1}{2} \\ \text{ and} \\ \forall y: \ \Pr_{z''}[S(y,z'')=1] \geq \frac{1}{2} \end{cases} \qquad \implies \Pr_{y,z''}[S(y,z'')=1] > \frac{1}{2}\,.$$

The reduction witnessing $\forall_q L \in \mathsf{C_=P}$ just outputs S, which has y, z'' as the random input bits.

Now we prove that $\vee L \in \mathsf{C_=P}$. Given $x_1,\ldots,x_\ell$, we run the reduction witnessing $L \in \mathsf{C_=P}$ on each x_i to get circuits $C_i(z)$ such that $x_i \in L \iff \Pr_z[C_i(z)=1] = 1/2$. We define $C_i^{(0)} = C_i$ and let $C_i^{(1)}$ be the same as C_i but with the output negated. Then we have

$$(x_1,\ldots,x_\ell) \in \vee L \iff \textstyle\prod_{i=1}^{\ell}\big(\Pr_z\big[C_i^{(0)}(z)=1\big] - \Pr_z\big[C_i^{(1)}(z)=1\big]\big) = 0\,.$$

We construct a new circuit $A(z')$ that satisfies

$$\Pr_{z'}[A(z')=1] \;=\; \frac{1}{2^{\ell-1}}\cdot \sum_{\substack{\text{even parity} \\ b\in\{0,1\}^\ell}} \prod_{i=1}^{\ell} \Pr_z\big[C_i^{(b_i)}(z)=1\big]$$

by randomly choosing an even parity b then running each $C_i^{(b_i)}$ independently. We also construct a new circuit $B(z')$ that satisfies

$$\Pr_{z'}[B(z')=1] \;=\; \frac{1}{2^{\ell-1}}\cdot \sum_{\substack{\text{odd parity} \\ b\in\{0,1\}^\ell}} \prod_{i=1}^{\ell} \Pr_z\big[C_i^{(b_i)}(z)=1\big]$$

by randomly choosing an odd parity b then running each $C_i^{(b_i)}$ independently. This implies that

$$(x_1,\ldots,x_\ell) \in \vee L \iff \Pr_{z'}[A(z')=1] - \Pr_{z'}[B(z')=1] = 0\,.$$

Now we construct a circuit $S(z'')$ such that

$$\Pr_{z''}[S(z'')=1] \;=\; \frac{1}{2}\cdot\Big(\Pr_{z'}[A(z')=1]+(1-\Pr_{z'}[B(z')=1])\Big)$$

by randomly choosing either A or B to run, and negating the output if B was chosen. This implies that $(x_1,\dots,x_\ell)\in\vee L \iff \Pr_{z''}[S(z'')=1]=1/2$. The reduction witnessing $\vee L\in\mathsf{C_{=}P}$ just outputs S.

A.2 The element distinctness problem

We supply a folklore proof of Lemma 4.6, which is used in the proof of Theorem 2.13. Consider the following algorithm.

```
   Input: x_1, x_2, ..., x_n ∈ {1, 2, ..., M}
   Output: are x_1, x_2, ..., x_n distinct?

1  execute the following two while loop computations in parallel until one of them halts:
2  while true do
3  |  choose a hash function f : {1, ..., M} → {1, ..., n} pairwise independently
4  |  if the number of pairs i < j such that f(x_i) = f(x_j) is ≤ 2n then
5  |  |  if ∃(i < j) such that f(x_i) = f(x_j) and x_i = x_j then halt and output "no" else halt and
   |  |  output "yes"
6  |  end
7  end
8  while true do
9  |  choose a pair of indices i < j uniformly at random
10 |  if x_i = x_j then halt and output "no"
11 end
```

Note that the algorithm never outputs an incorrect answer. If the number of pairs $i<j$ such that $x_i=x_j$ is $\geq n/2$ then the second **while** loop halts after $\leq \binom{n}{2}/(n/2)=n-1$ iterations in expectation. Otherwise, the expectation (over the choice of f) of the number of pairs $i<j$ such that $f(x_i)=f(x_j)$ is $<(n/2)+\binom{n}{2}\cdot(1/n)<n$, and so the first **while** loop has $>1/2$ probability of halting in each iteration and thus halts after <2 iterations in expectation. Line 4 takes linear time by building lists of input elements that hash to each bucket, and Line 5 takes linear time by brute force (or sorting) on each bucket, so the first **while** loop would take linear time in expectation.

References

[1] DAVID J. ALDOUS: More uses of exchangeability: Representations of complex random structures. In *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*, pp. 35–63. Cambridge Univ. Press, 2010. [arXiv:0909.4339] 2

[2] NOGA ALON, ALEXANDR ANDONI, TALI KAUFMAN, KEVIN MATULEF, RONITT RUBINFELD, AND NING XIE: Testing k-wise and almost k-wise independence. In *Proc. 39th STOC*, pp. 496–505. ACM Press, 2007. [doi:10.1145/1250790.1250863] 2

[3] SANJEEV ARORA AND BOAZ BARAK: *Computational Complexity – A Modern Approach*. Cambridge Univ. Press, 2009. 4

[4] TUĞKAN BATU, SANJOY DASGUPTA, RAVI KUMAR, AND RONITT RUBINFELD: The complexity of approximating the entropy. *SIAM J. Comput.*, 35(1):132–150, 2005. Preliminary version in CCC'02. [doi:10.1137/S0097539702403645] 2

[5] TUĞKAN BATU, ELDAR FISCHER, LANCE FORTNOW, RAVI KUMAR, RONITT RUBINFELD, AND PATRICK WHITE: Testing random variables for independence and identity. In *Proc. 42nd FOCS*, pp. 442–451. IEEE Comp. Soc. Press, 2001. [doi:10.1109/SFCS.2001.959920] 2, 21

[6] TUĞKAN BATU, LANCE FORTNOW, RONITT RUBINFELD, WARREN SMITH, AND PATRICK WHITE: Testing closeness of discrete distributions. *J. ACM*, 4, 2013. Preliminary version in FOCS'00. [doi:10.1145/2432622.2432626, arXiv:1009.5397] 2

[7] TUĞKAN BATU, RAVI KUMAR, AND RONITT RUBINFELD: Sublinear algorithms for testing monotone and unimodal distributions. In *Proc. 36th STOC*, pp. 381–390. ACM Press, 2004. [doi:10.1145/1007352.1007414] 2

[8] CHRISTOPHER BECK, RUSSELL IMPAGLIAZZO, AND SHACHAR LOVETT: Large deviation bounds for decision trees and sampling lower bounds for AC^0-circuits. In *Proc. 53rd FOCS*, pp. 101–110. IEEE Comp. Soc. Press, 2012. [doi:10.1109/FOCS.2012.82] 3

[9] ANDRÉ BERTHIAUME AND GILLES BRASSARD: The quantum challenge to structural complexity theory. In *Proc. 7th IEEE Conf. on Structure in Complexity Theory (SCT'92)*, pp. 132–137, 1992. [doi:10.1109/SCT.1992.215388] 24

[10] ANDREJ BOGDANOV, ELCHANAN MOSSEL, AND SALIL P. VADHAN: The complexity of distinguishing Markov random fields. In *Proc. 12th Internat. Workshop on Randomization and Computation (RANDOM'08)*, volume 5171, pp. 331–342, 2008. [doi:10.1007/978-3-540-85363-3_27] 3

[11] BERND BORCHERT, LANE A. HEMASPAANDRA, AND JÖRG ROTHE: Restrictive acceptance suffices for equivalence problems. *LMS Journal of Computation and Mathematics*, 3:86–95, 2000. Preliminary version in FCT'99. [doi:10.1112/S146115700000022X] 2

[12] BERND BORCHERT AND FRANK STEPHAN: Looking for an analogue of Rice's Theorem in circuit complexity theory. *Mathematical Logic Quarterly*, 46(4):489–504, 2000. Preliminary versions in KGC'97 and ECCC. [doi:10.1002/1521-3870(200010)46:4<489::AID-MALQ489>3.0.CO;2-F] 24

[13] SIU ON CHAN, ILIAS DIAKONIKOLAS, GREGORY VALIANT, AND PAUL VALIANT: Optimal algorithms for testing closeness of discrete distributions. In *Proc. 25th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'14)*, volume 6, pp. 1193–1203. ACM Press, 2014. [doi:10.1137/1.9781611973402.88] 2

[14] CONSTANTINOS DASKALAKIS, ILIAS DIAKONIKOLAS, ROCCO SERVEDIO, GREGORY VALIANT, AND PAUL VALIANT: Testing k-modal distributions: Optimal algorithms via reductions. In *Proc. 24th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'13)*, pp. 1833–1852. ACM Press, 2013. [doi:10.1137/1.9781611973105.131] 2

[15] ANINDYA DE AND THOMAS WATSON: Extractors and lower bounds for locally samplable sources. *ACM Transactions on Computation Theory*, 4(3), 2012. Preliminary versions in RANDOM'11 and ECCC. [doi:10.1145/2141938.2141941] 3

[16] PERCI DIACONIS AND DAVID FREEDMAN: Finite exchangeable sequences. *Ann. Probab.*, 8(4):745–764, 1980. [doi:10.1214/aop/1176994663] 2

[17] ZEEV DVIR, DAN GUTFREUND, GUY ROTHBLUM, AND SALIL P. VADHAN: On approximating the entropy of polynomial mappings. In *Proc. 2nd Innovations in Computer Science Conf. (ICS'11)*, pp. 460–475, 2011. Preliminary version in ECCC. 3

[18] STEPHEN FENNER, FREDERIC GREEN, STEVEN HOMER, AND RANDALL PRUIM: Quantum NP is hard for PH. In *Proc. 6th Italian Conf. on Theoretical Computer Science*, pp. 241–252, 1998. Expanded version in ECCC. 2

[19] ODED GOLDREICH: *Computational Complexity – A Conceptual Perspective*. Cambridge Univ. Press, 2008. 4

[20] ODED GOLDREICH, AMIT SAHAI, AND SALIL P. VADHAN: Can statistical zero knowledge be made non-interactive? or On the relationship of SZK and NISZK. In *Proc. 19th Internat. Cryptology Conf.*, pp. 467–484, 1999. Preliminary version in ECCC. [doi:10.1007/3-540-48405-1_30] 2, 3

[21] ODED GOLDREICH AND SALIL P. VADHAN: Comparing entropies in statistical zero-knowledge with applications to the structure of SZK. In *Proc. 14th IEEE Conf. on Computational Complexity (CCC'99)*, pp. 54–73, 1999. [doi:10.1109/CCC.1999.766262] 2

[22] ODED GOLDREICH AND SALIL P. VADHAN: On the complexity of computational problems regarding distributions. In *Studies in Complexity and Cryptography*, volume 6650 of *LNCS*, pp. 390–405. Springer, 2011. Also available at ECCC. [doi:10.1007/978-3-642-22670-0_27] 2

[23] RUSSELL IMPAGLIAZZO AND AVI WIGDERSON: $\mathrm{P} = \mathrm{BPP}$ if E requires exponential circuits: Derandomizing the XOR Lemma. In *Proc. 29th STOC*, pp. 220–229. ACM Press, 1997. [doi:10.1145/258533.258590] 26

[24] JESSE KAMP, ANUP RAO, SALIL P. VADHAN, AND DAVID ZUCKERMAN: Deterministic extractors for small-space sources. *J. Comput. System Sci.*, 77(1):191–220, 2011. Preliminary version in STOC'06. [doi:10.1016/j.jcss.2010.06.014] 3

[25] JOHN KINGMAN: Uses of exchangeability. *Ann. Probab.*, 6(2):183–197, 1978. 2

[26] REUT LEVI, DANA RON, AND RONITT RUBINFELD: Testing properties of collections of distributions. *Theory of Computing*, 9:295–347, 2013. Preliminary versions in ICS'10 and ECCC. [doi:10.4086/toc.2013.v009a008] 2

[27] SHACHAR LOVETT AND EMANUELE VIOLA: Bounded-depth circuits cannot sample good codes. *Comput. Complexity*, 21(2):245–266, 2012. Preliminary versions in CCC'12 and ECCC. [doi:10.1007/s00037-012-0039-3] 3

[28] SOFYA RASKHODNIKOVA, DANA RON, AMIR SHPILKA, AND ADAM SMITH: Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM J. Comput.*, 39(3):813–842, 2009. Preliminary version in FOCS'07. [doi:10.1137/070701649] 2

[29] RONITT RUBINFELD AND ROCCO SERVEDIO: Testing monotone high-dimensional distributions. *Random Structures Algorithms*, 34(1):24–44, 2009. Preliminary version in STOC'05. [doi:10.1002/rsa.20247] 2

[30] RONITT RUBINFELD AND NING XIE: Testing non-uniform k-wise independent distributions over product spaces. In *Proc. 37th Internat. Colloq. on Automata, Languages and Programming (ICALP'10)*, pp. 565–581, 2010. [doi:10.1007/978-3-642-14165-2_48] 2

[31] AMIT SAHAI AND SALIL P. VADHAN: A complete problem for statistical zero knowledge. *J. ACM*, 50(2):196–249, 2003. Preliminary versions in FOCS'97, ECCC and Cryptology ePrint Archive. [doi:10.1145/636865.636868] 2, 3, 7, 20

[32] JUN TARUI: Probabilistic polynomials, AC^0 functions, and the polynomial-time hierarchy. *Theoret. Comput. Sci.*, 113(1):167–183, 1993. Preliminary version in STACS'91. [doi:10.1016/0304-3975(93)90214-E] 2

[33] SEINOSUKE TODA AND MITSUNORI OGIWARA: Counting classes are at least as hard as the polynomial-time hierarchy. *SIAM J. Comput.*, 21(2):316–328, 1992. Preliminary version in CCC'91. [doi:10.1137/0221023] 2

[34] LUCA TREVISAN, SALIL P. VADHAN, AND DAVID ZUCKERMAN: Compression of samplable sources. *Comput. Complexity*, 14(3):186–227, 2005. Preliminary versions in CCC'04 and ECCC. [doi:10.1007/s00037-005-0198-6] 3

[35] PAUL VALIANT: Testing symmetric properties of distributions. *SIAM J. Comput.*, 40(6):1927–1968, 2011. Preliminary versions in STOC'08 and ECCC. [doi:10.1137/080734066] 2

[36] EMANUELE VIOLA: The complexity of distributions. *SIAM J. Comput.*, 41(1):191–218, 2012. Preliminary version in FOCS'10. [doi:10.1137/100814998] 3

[37] EMANUELE VIOLA: Extractors for circuit sources. *SIAM J. Comput.*, 43(2):655–672, 2014. Preliminary versions in FOCS'11 and ECCC. [doi:10.1137/11085983X] 3

[38] KLAUS W. WAGNER: The complexity of combinatorial problems with succinct input representation. *Acta Informatica*, 23(3):325–356, 1986. [doi:10.1007/BF00289117] 2

[39] THOMAS WATSON: The complexity of deciding statistical properties of samplable distributions. In *Proc. 31st Symp. Theoretical Aspects of Comp. Sci. (STACS'14)*, pp. 663–674, 2014. (Preliminary version of this paper). [doi:10.4230/LIPIcs.STACS.2014.663] 1

[40] THOMAS WATSON: The complexity of estimating min-entropy. *Comput. Complexity*, 2014 (online). Preliminary version in ECCC. [doi:10.1007/s00037-014-0091-2] 2, 3

AUTHOR

Thomas Watson
Postdoctoral researcher
University of Toronto, Toronto, ON
thomasw@cs.toronto.edu
`http://www.cs.toronto.edu/~thomasw/`

ABOUT THE AUTHOR

THOMAS WATSON is a postdoc at the University of Toronto, supervised by Toniann Pitassi. He received his Ph. D. in computer science from the University of California, Berkeley, supervised by Luca Trevisan. He received his undergraduate degree in computer science, math, and computer engineering from the University of Wisconsin-Madison, where his interest in theoretical computer science research was sparked and fostered by Dieter van Melkebeek. His research interests include most of complexity theory, but recently he has been working on communication complexity.