Electronic Colloquium on Computational Complexity, Report No. 124 (2013)

# The Complexity of Deciding Statistical Properties of Samplable Distributions

Thomas Watson[*]

September 8, 2013

## Abstract

We consider the problems of deciding whether the joint distribution sampled by a given circuit satisfies 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 $\mathrm{C}_=\mathrm{P}$-complete, by showing that the following promise problem (which is a restriction of all the above problems) is $\mathrm{C}_=\mathrm{P}$-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 introduce a bounded-error version of $\mathrm{C}_=\mathrm{P}$, which we call $\mathrm{BC}_=\mathrm{P}$, and we investigate its structural properties.

## 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 [BFR$^+$13, BFF$^+$01, BDKR05, BKR04, RS09, AAK$^+$07, RRSS09, Val11, RX10, LRR13, DDS$^+$13, CDVV13].

The other viewpoint is the non-black-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

[*] Department of Computer Science, University of Toronto.

ISSN 1433-8092

set.) For arbitrary circuit samplers, the papers [SV03, GSV99, Wat12] contain completeness results for various approximation problems concerning statistical distance, Shannon entropy, and min-entropy. See [GV11] for a survey of both the black-box and the non-black-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 [Kin78, Ald10], 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] [DF80]. 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 no harder than any of those problems, are complete for the complexity class $\mathrm{C_{=}P}$. This class was introduced in [Wag86] as part of the counting hierarchy, and it can be viewed as a class that captures "exact counting" of NP witnesses. The class $\mathrm{C_{=}P}$ is at least as hard as the polynomial-time hierarchy, since $\mathrm{PH} \subseteq \mathrm{BP} \cdot \mathrm{C_{=}P}$ [TO92] and even $\mathrm{PH} \subseteq \mathrm{ZP} \cdot \mathrm{C_{=}P}$ [Tar93]. It is no harder than "threshold counting", since $\mathrm{C_{=}P} \subseteq \mathrm{PP}$, but neither is it substantially easier, since $\mathrm{PP} \subseteq \mathrm{NP}^{\mathrm{C_{=}P}}$. The class $\mathrm{C_{=}P}$ has been given several names and characterizations; it equals the classes coNQP [FGHP98] and ES [BHR99].

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 measures are parallel time and memory space. One model of efficient parallel time computation is $\mathrm{AC}^0$ (constant-depth unbounded fan-in circuits with AND, OR, and NOT gates). Papers that study $\mathrm{AC}^0$ circuits that sample distributions include [Vio12, LV12, Vio11, BIL12]. 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 [Vio12, DW12, DGRV11, Wat12] as well as a large collection of papers investigating the possibility of implementing pseudorandom generators locally. (See [DW12] 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 [TVZ05], extraction [KRVZ11], and min-entropy estimation [Wat12]. One more paper worth mentioning is [BMV08], which considers Markov random fields as succinct descriptions of distributions (though these descriptions would not be considered "samplers").

We prove that our $\mathrm{C_{=}P}$-completeness results hold even when restricted to samplers that are $\mathrm{AC}^0$-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

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

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.

In this paper we also consider a "bounded-error" version of $\mathrm{C_{=}P}$, which we call $\mathrm{BC_{=}P}$ 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 $i^{\text{th}}$ 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 $\leq s$ 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 deterministic polynomial-time mapping reductions.

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

### 2.1 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 nonnegative integers). Uncorrelated is the same as independent if $k = 1$. We consider the following extreme notion of a distribution being nonuniform.

**Definition 1.** *A joint distribution is* discordant *if there are $\geq 2$ coordinates and every pair of coordinates is neither uncorrelated nor identically distributed.*

**Definition 2.** Panoptic-Stats *is the following promise problem.*

$$\begin{aligned} \textsc{Panoptic-Stats}_{\text{YES}} &= \big\{S \;:\; S(U) \textit{ is uniform}\big\} \\ \textsc{Panoptic-Stats}_{\text{NO}} &= \big\{S \;:\; S(U) \textit{ is discordant}\big\} \end{aligned}$$

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

**Fact 1.** 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 1 are complete for the complexity class $\mathrm{C_{=}P}$. In fact, the $\mathrm{C_{=}P}$-hardness of each of the individual languages in Fact 1 is fairly simple to prove, but the $\mathrm{C_{=}P}$-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 $\mathrm{C_{=}P}$-hardness of all the languages in Fact 1.

To prove the $\mathrm{C_{=}P}$-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 3.** *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 4.** 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 $m$.*

$\mathrm{prC_{=}P}$ is the class of promise problems for which there exists a polynomial-time randomized algorithm $M$ that accepts with probability $\frac{1}{2}$ on YES instances, and accepts with probability $\neq \frac{1}{2}$ on NO instances. Here we use a standard model of computation in which randomized algorithms have access to independent unbiased coin flips. We use the following equivalent definition of $\mathrm{prC_{=}P}$.

**Definition 5.** $\mathrm{prC_{=}P}$ *is the class of all promise problems reducible to the following promise problem* Uniform-Bit*.*

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

$\mathrm{C_{=}P}$ *is defined as the class of languages in* $\mathrm{prC_{=}P}$*.*

**Theorem 1.** *For every polite $\kappa, \nu, \sigma$ with $\kappa\nu \leq o(\sigma)$,* Panoptic-Stats$^{\kappa,\nu,\sigma}$ *is* $\mathrm{prC_{=}P}$*-hard.*

**Theorem 2.** *For every polite $\kappa, \nu, \sigma$ with $\kappa\nu + \nu^2 \leq o(\sigma)$,* Panoptic-Stats$^{\kappa,\nu,\sigma}$ *is* $\mathrm{prC_{=}P}$*-hard even when restricted to samplers that are* $\mathrm{AC}^0$*-type circuits with depth* 3 *and top fan-in* 2*.*

**Theorem 3.** *All the languages listed in Fact 1 are in* $\mathrm{C_{=}P}$*.*

Consequently, all the languages listed in Fact 1 are $\mathrm{C_{=}P}$-complete, even when restricted to $(\kappa, \nu, \sigma)$-samplers (like in Definition 4) with polite $\kappa, \nu, \sigma$ 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 $\mathrm{C}_=\mathrm{P}$-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 1 also shows that this problem is $\mathrm{C}_=\mathrm{P}$-hard. Both these problems are in $\mathrm{C}_=\mathrm{P}$ by the same techniques used in the proof of Theorem 3.

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

**Theorem 5.** *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 6.** *There exists a linear-time 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.

## 3 Completeness Results

We prove a key lemma in Section 3.1. Then we use the key lemma to prove Theorems 1 and 2 in Section 3.2. Then we prove Theorem 3 in Section 3.3.

### 3.1 The Key Lemma

The following is the key lemma in the proof of Theorem 1. It can be interpreted qualitatively as a certain type of amplification.

**Lemma 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.*

$$\begin{aligned} S(U) \text{ is uniform} \quad &\Longrightarrow \quad T(U) \text{ is uniform} \\ S(U) \text{ is nonuniform} \quad &\Longrightarrow \quad T(U) \text{ is discordant} \end{aligned}$$

*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 $\ell^*$ be the least such $\ell$
    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 $\Pr[S(U) = 1] = p \neq \frac{1}{2}$. 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 $\ell^*$ does not exist then we define $\ell^*$ to be an arbitrary value $> n$.

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

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

Similarly, $\Pr[D_j = 1] = p \cdot \frac{1}{2^{j-1}} + \frac{1}{2} \cdot \big(1 - \frac{1}{2^{j-1}}\big)$. Since $p \neq \frac{1}{2}$, and since $\Pr[D_i = 1]$ and $\Pr[D_j = 1]$ are different convex combinations of $p$ and $\frac{1}{2}$, that means they are not equal. More formally,

$$\Pr[D_i = 1] - \Pr[D_j = 1] \;=\; \big(p - \tfrac{1}{2}\big)\big(\tfrac{1}{2^{i-1}} - \tfrac{1}{2^{j-1}}\big) \;\neq\; 0.$$

On the other hand, suppose $i = 1$. Then $\Pr[D_i = 1] = \frac{1}{2}$, and $\Pr[D_j = 1]$ is a nontrivial convex combination of $p$ and $\frac{1}{2}$ 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 $\Pr\big[D_j = 1 \;\big|\; D_i = 1\big] = \frac{1}{2}$, and

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

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

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

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

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

It follows that

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

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

**Lemma 2.** *Lemma 1 holds even when $T$ is required to be an $\mathrm{AC}^0$-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 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 $\frac{1}{2}$ iff $F$ accepts a uniformly random input with probability $\frac{1}{2^q}$. Let $x_1, x_2, \ldots, x_r$ denote the input bits of $F$. Construct a new CNF $F'$ with input bits $x_0, x_1, \ldots, x_r$ 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

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

it follows that $F$ accepts with probability $\frac{1}{2^q}$ iff $F'$ accepts with probability $\frac{1}{2}$. Now to implement $T$, we include a copy of $F'$ as well as the random input bits $a_1, a_2, \ldots, a_n$. The 1st 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. □

### 3.2 $\mathrm{prC}_=\mathrm{P}$-Hardness

We need one final corollary before we are ready to put the pieces together to prove Theorems 1 and 2.

**Corollary 1.** *Lemmas 1 and 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 1) or $O(kn + n^2 + s)$ (for Lemma 2).*

*Proof.* If $T'$ is the output of the algorithm from Lemma 1 or 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

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

□

*Proof of Theorem 1.* We reduce Uniform-Bit to Panoptic-Stats$^{\kappa,\nu,\sigma}$. Let $c$ be the constant factor in the big $O$ in Corollary 1. 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 1 (based on Lemma 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 1 runs in time $O(kn + s)$. □

*Proof of Theorem 2.* We reduce Uniform-Bit to Panoptic-Stats$^{\kappa,\nu,\sigma}$ restricted as in the statement of Theorem 2. Let $c$ be the constant factor in the big $O$ in Corollary 1. 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 1 (based on Lemma 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 1 runs in time $O(kn + n^2 + s)$. □

### 3.3 Containment in $\mathrm{C_{=}P}$

In the proof of Theorem 3 we use the following lemma, which states that $\mathrm{C_{=}P}$ is closed under exponential conjunctions and polynomial disjunctions. We supply a folklore proof of this lemma in Appendix A.1.

**Lemma 3.** *If $L \in \mathrm{C_{=}P}$ then both of the following hold.*

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

*Proof of Theorem 3.* 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 $\in \mathrm{C_{=}P}$. 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 it suffices to show that $L \in \mathrm{C_{=}P}$. 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 $i^{\text{th}}$ coordinate of the output of the $i^{\text{th}}$ run is $y_i$.

Now we prove that $K$-WISE-EXCHANGEABLE $\in \mathrm{C_{=}P}$. 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 $I \subseteq \{1,\ldots,n\}$ of size $\min(K,n)$, a permutation $\pi$ on $\{1,\ldots,\min(K,n)\}$, 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 $\pi$. Thus by Lemma 3 it suffices to show that $L \in \mathrm{C_{=}P}$. 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 $\in \mathrm{C_{=}P}$. Note that if we define the language $L = \{(S,i,j) \ : \ S(U)_i \text{ and } S(U)_j \text{ are uncorrelated}\}$, then 2-EXISTS-UNCORRELATED reduces to $\vee L$ by mapping a $(k,n)$-sampler $S$ to $\big((S,1,2),(S,1,3),(S,1,4),\ldots,(S,n{-}1,n)\big)$. Thus by Lemma 3 it suffices to show that $L \in \mathrm{C_{=}P}$. 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 $\frac{1}{2^{2k}} \cdot y_i \cdot y_j$ so that

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

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

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

# 4 Algorithmic Results

We prove Theorems 4, 5, 6 in Sections 4.1, 4.2, 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 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 Proof of Theorem 4

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 $\ell$ 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 $\frac{1}{2^\ell}$ (since they only depend on $\ell$ input bits), but the product of the marginal probabilities that each edge on the cycle evaluates to 1 must be either $\frac{1}{2^{\ell+1}}$ (if there is an AND-type edge) or $\frac{3}{2^{\ell+1}}$ (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 $\ell$ 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 $\frac{1}{2^\ell}$ (since they only depend on $\ell$ input bits), but the product of the marginal probabilities that each edge on the path evaluates to 1 must be either $\frac{1}{2^{\ell+1}}$ (if neither $e_1$ nor $e_2$ is OR-type) or $\frac{3}{2^{\ell+1}}$ (if exactly one of $e_1, e_2$ is OR-type) or $\frac{9}{2^{\ell+1}}$ (if both $e_1$ and $e_2$ are OR-type). Hence $S(U)$ is not fully independent.

## 4.2 Proof of Theorem 5

We begin with a lemma.

**Lemma 4.** *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.* 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

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

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 $h \in \{1, \ldots, n\}$), let $E$ denote the event that $D_h = b_h$ for all $h \in \{1, \ldots, n\}$, and let $E'$ denote the event that $D'_h = b_h$ for all $h \in \{1, \ldots, n\}$. 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

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

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

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

This finishes the proof of Lemma 4. □

Now we present the proof of Theorem 5. Consider a 2-local $(1, n)$-sampler $S$. If condition (1) from Lemma 4 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.[2]

(iii) The graph is a triangle, and there is agreement at all nodes.[3]

(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 $\frac{1}{4}$ 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 have probability $\frac{1}{8}$, 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 $1 - \frac{n}{4}$ 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 $\frac{1}{4}$ and all 0's with probability $\frac{3}{4}$, 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 5.** *If none of (i)–(vii) hold then at least one of (A)–(E) holds.*

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

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

*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. □

**Lemma 6.** *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 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 $\frac{1}{3}$. If (A) holds and $e_1, e_3$ share only one endpoint and $e_1, e_2$ are disjoint then it goes either from $\frac{1}{2}$ to $\frac{1}{6}$ (if $e_1, e_3$ agree) or from 0 to $\frac{1}{3}$. If (A) holds and $e_3, e_1, e_2$ form a simple path then it goes either from $\frac{1}{2}$ to 0 (if $e_1, e_3$ agree and $e_1, e_2$ agree) or from $\frac{1}{2}$ to $\frac{1}{4}$ (if $e_1, e_3$ agree and $e_1, e_2$ disagree) or from 0 to $\frac{1}{2}$ (if $e_1, e_3$ disagree and $e_1, e_2$ agree) or from 0 to $\frac{1}{4}$.

If (B) holds then it goes either from $\frac{1}{2}$ to 0 (if $e_1, e_3$ agree) or from 0 to $\frac{1}{2}$. 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 $\frac{1}{2}$ to 0. If (E) holds then it goes from 1 to 0. □

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.[4]

(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 4. 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 $\frac{1}{2}$ and all 0's with probability $\frac{1}{2}$, 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.

(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 7.** *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

[4] 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.

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 $\geq 1$). Thus at least one of $e$'s nodes is not on $P$, so (B) holds. □

**Lemma 8.** *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 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. □

**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 Proof of Theorem 6

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.[5] We supply a folklore proof of the following lemma in Appendix A.2.

**Lemma 9.** *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 4 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 BC$_=$P

We now consider the bounded-error version of C$_=$P, which does not seem to have been defined or studied in the literature before.[6]

**Definition 6.** prBC$_=$P *is the class of all promise problems reducible to the following promise problem* Bounded-Uniform-Bit.

$$\begin{aligned} \textsc{Bounded-Uniform-Bit}_{\text{YES}} &= \big\{S \;:\; S \textit{ is a } (1,1)\textit{-sampler and } S(U) \textit{ is uniform}\big\} \\ \textsc{Bounded-Uniform-Bit}_{\text{NO}} &= \big\{S \;:\; S \textit{ is a } (1,1)\textit{-sampler and } \big|\Pr[S(U)=1]-\tfrac{1}{2}\big| \geq \tfrac{1}{4}\big\} \end{aligned}$$

BC$_=$P *is defined as the class of languages in* prBC$_=$P.

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

**Lemma 10.** *For all languages $L$, the following are equivalent.*

*(1) For some polynomial $q$, there is a reduction that takes $x$ and outputs a $(1,1)$-sampler $S$ such that the following both hold.*[7]

$$\begin{aligned} x \in L &\implies S(U) \textit{ is uniform} \\ x \notin L &\implies \big|\Pr[S(U)=1]-\tfrac{1}{2}\big| \geq \tfrac{1}{q(|x|)} \end{aligned}$$

*(2) $L \in$ BC$_=$P.*

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

[6] A more extreme version, in which YES instances have acceptance probability $\frac{1}{2}$ and NO instances have acceptance probability 0, has been studied before and is known as C$_{==}$P[half] [BB92] and HalfP [BS00].

[7] 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 reduction that takes $x$ and outputs a $(1,1)$-sampler $S$ such that the following both hold.*

$$\begin{aligned} x \in L &\implies S(U) \textit{ is uniform} \\ x \notin L &\implies \Pr[S(U)=1] \le \tfrac{1}{2^{Q(|x|)}} \end{aligned}$$

*Proof.* Clearly (3) ⇒ (2) ⇒ (1), so we just need to demonstrate (1) ⇒ (3). Assume (1). By the standard trick for making $\mathrm{C_{=}P}$ have "1-sided error" (see Appendix A.1), it follows that there is a similar reduction mapping $x$ to some (different) $(1,1)$-sampler $S$ that achieves $\Pr[S(U)=1] \le \frac{1}{2} - \frac{1}{q(|x|)^2}$ in the case $x \notin L$. Construct a new circuit $S'$ that performs the following computation.

```
let $m = 2q(|x|)^4 Q(|x|)$ and $t = \big(\frac{1}{2} - \frac{1}{2q(|x|)^2}\big) \cdot m$
choose a uniformly random bit $b$
**if** $b = 0$ **then** run $S$ independently $m$ times, and accept iff $\ge t$ 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 hardwired into $S'$. If $S$ accepts with probability $\frac{1}{2}$ then the probability that $\ge t$ 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 $\frac{1}{2}$. If $S$ accepts with probability $\le \frac{1}{2} - \frac{1}{q(|x|)^2}$ then by a standard concentration bound, the probability that $\ge t$ of $m$ runs of $S$ accept is $\le 2^{-m/2q(|x|)^4} = \frac{1}{2^{Q(|x|)}}$. Also by a standard concentration bound, $\big(1 - \frac{1}{2^m}\sum_{i=\lceil t\rceil}^{m}\binom{m}{i}\big) \le 2^{-m/2q(|x|)^4} = \frac{1}{2^{Q(|x|)}}$. Hence if $x \notin L$ then $S'$ accepts with probability $\le \frac{1}{2}\cdot\frac{1}{2^{Q(|x|)}} + \frac{1}{2}\cdot\frac{1}{2^{Q(|x|)}} = \frac{1}{2^{Q(|x|)}}$. The reduction for (3) just outputs $S'$. □

Note that $\mathrm{coRP} \subseteq \mathrm{BC_{=}P}$. The 1-sided error property in part (3) of Lemma 10 implies that $\mathrm{BC_{=}P} \subseteq \mathrm{BPP}$. Thus $\mathrm{BC_{=}P}$ is presumably closed under complement (since presumably $\mathrm{P} = \mathrm{BC_{=}P} = \mathrm{BPP}$ [IW97]), but proving this seems out of reach. We now apply Lemma 10 to prove that $\mathrm{BC_{=}P}$ is closed under union, intersection, disjunction (i.e., $\vee L \in \mathrm{BC_{=}P}$ if $L \in \mathrm{BC_{=}P}$, where $\vee L = \big\{(x_1,\ldots,x_\ell) \,:\, x_i \in L \text{ for some } i\big\}$), and conjunction (i.e., $\wedge L \in \mathrm{BC_{=}P}$ if $L \in \mathrm{BC_{=}P}$, where $\wedge L = \big\{(x_1,\ldots,x_\ell) \,:\, x_i \in L \text{ for all } i\big\}$). The proof of Lemma 3 in Appendix A.1 showing that $\mathrm{C_{=}P}$ is closed under disjunction does not work to show that $\mathrm{BC_{=}P}$ is closed under disjunction.

**Theorem 7.** $\mathrm{BC_{=}P}$ *is closed under disjunction.*

*Proof.* Assuming $L \in \mathrm{BC_{=}P}$, we exhibit a reduction witnessing $\vee L \in \mathrm{BC_{=}P}$. 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 $n \ge \ell$. By (2) ⇒ (3) in Lemma 10, there is a reduction that takes $x_i$ and outputs $C_i$ such that the following both hold.

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

Our reduction witnessing $\vee L \in \mathrm{BC_{=}P}$ 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 $(x_1, \dots, x_\ell) \in \vee L$ then $S$ is taking the parity of $\ell$ independent bits at least one of which is uniform, so $S(U)$ is uniform. If $(x_1, \dots, x_\ell) \notin \vee L$ then letting $z_1, \dots, z_\ell$ denote $S$'s random input, we have

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

□

**Theorem 8.** BC$_=$P *is closed under conjunction.*

*Proof.* Assuming $L \in \mathrm{BC_=P}$, we exhibit a reduction witnessing $\wedge L \in \mathrm{BC_=P}$. We are given $(x_1, \dots, x_\ell)$. Lemma 10 implies that 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.

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

Our reduction witnessing $\wedge L \in \mathrm{BC_=P}$ runs the above reduction for $L$ on each $x_i$ to obtain the circuits $C_i$, then outputs a circuit $S$ that chooses $i \in \{1, \dots, \ell\}$ 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 $(x_1, \dots, x_\ell) \in \wedge L$ then $S(U)$ is uniform since $\Pr[C_i(U) = 1] = \frac{1}{2}$ for all $i$. If $(x_1, \dots, x_\ell) \notin \wedge L$ 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) ⇒ (2) in Lemma 10, $\wedge L \in \mathrm{BC_=P}$. □

**Corollary 2.** BC$_=$P *is closed under union and intersection.*

*Proof.* This follows by the same proof techniques used for Theorems 7 and 8. Alternatively, this follows in a black-box fashion from Theorems 7 and 8: Assuming $L_1 \in \mathrm{BC_=P}$ and $L_2 \in \mathrm{BC_=P}$, we also have $L \in \mathrm{BC_=P}$ where $L = \big\{(x, i) \ : \ i \in \{1, 2\} \text{ and } x \in L_i\big\}$, and thus $\vee L \in \mathrm{BC_=P}$ and $\wedge L \in \mathrm{BC_=P}$. 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 $L_1 \cup L_2 \in \mathrm{BC_=P}$ and $L_1 \cap L_2 \in \mathrm{BC_=P}$. □

# 6 Open Problems

There are plenty of open problems concerning the complexity of deciding statistical properties of joint distributions with low-complexity samplers. What can be said about depth-2 circuits? What about $d$-local samplers when $d > 2$ or $k > 1$? What are the complexities of deciding full independence and exchangeability for logspace samplers?

For general circuit samplers, what is the complexity of deciding whether a distribution is a mixture of i.i.d. distributions? It is not clear whether this problem is in 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 C$_=$P.

Can completeness results be proven for approximate versions of the problems we considered in this paper? Goldreich, Sahai, and Vadhan [GSV99] proved that the promise problem of deciding whether a sampler's distribution is statistically close or far from uniform is prNISZK-complete. What about analogous problems where "uniform" is replaced by "fully independent" or "exchangeable" or "pairwise independent"?

Are there other interesting structural properties or applications of BC$_=$P?

# A Folklore Proofs

We use this appendix to supply some folklore proofs.

## A.1 Closure of C$_=$P

We supply a folklore proof of Lemma 3, which is used in the proof of Theorem 3. Assume $L \in \mathrm{C_=P}$. First we prove that $\forall_q L \in \mathrm{C_=P}$.

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.

$$\begin{aligned} x \in \forall_q L \quad &\Longrightarrow \quad \forall y: \ \ \Pr_z[C(y,z)=1] = \tfrac{1}{2} \\ x \notin \forall_q L \quad &\Longrightarrow \quad \exists y: \ \ \Pr_z[C(y,z)=1] \neq \tfrac{1}{2} \end{aligned}$$

The circuit $C$ runs the reduction witnessing $L \in \mathrm{C_=P}$ 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

$$\Pr_{z'}[A(y,z')=1] \ = \ \tfrac{1}{2} \cdot \big(\Pr_z[C(y,z)=1]^2 + \Pr_z[\overline{C}(y,z)=1]^2\big)$$

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

$$\Pr_{z'}[B(y,z')=1] \ = \ \tfrac{1}{2} \cdot \big(2 \cdot \Pr_z[C(y,z)=1] \cdot \Pr_z[\overline{C}(y,z)=1]\big)$$

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

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

which implies the following.

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

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

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

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

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

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

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

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

Now we prove that $\vee L \in \mathrm{C_=P}$. Given $x_1, \ldots, x_\ell$, we run the reduction witnessing $L \in \mathrm{C_=P}$ on each $x_i$ to get circuits $C_i(z)$ such that $x_i \in L \iff \Pr_z[C_i(z)=1] = \frac{1}{2}$. 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

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

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

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

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

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

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

## A.2 The Element Distinctness Problem

We supply a folklore proof of Lemma 9, which is used in the proof of Theorem 6. Consider the following algorithm.

```
   Input: $x_1, x_2, \ldots, x_n \in \{1, 2, \ldots, M\}$
   Output: are $x_1, x_2, \ldots, x_n$ 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 $f : \{1, \ldots, M\} \to \{1, \ldots, n\}$ pairwise independently
4      if the number of pairs $i < j$ such that $f(x_i) = f(x_j)$ is $\leq 2n$ then
5          if $\exists (i < j)$ such that $f(x_i) = f(x_j)$ and $x_i = x_j$ then halt and output “no”
6          else halt and output “yes”
7      end
8  end
9  while true do
10     choose a pair of indices $i < j$ uniformly at random
11     if $x_i = x_j$ then halt and output “no”
12 end
```

Note that the algorithm never outputs an incorrect answer. If the number of pairs $i < j$ such that $x_i = x_j$ is $\geq \frac{n}{2}$ then the second while loop halts after $\leq \binom{n}{2} / \frac{n}{2} = n - 1$ 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 $< \frac{n}{2} + \binom{n}{2} \cdot \frac{1}{n} < n$, and so the first while loop has $> \frac{1}{2}$ 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

[AAK$^+$07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing $k$-wise and almost $k$-wise independence. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, pages 496–505, 2007.

[Ald10] David Aldous. More uses of exchangeability: Representations of complex random structures. In *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*, pages 35–63. Cambridge University Press, 2010.

[BB92] André Berthiaume and Gilles Brassard. The quantum challenge to structural complexity theory. In *Proceedings of the 7th Structure in Complexity Theory Conference*, pages 132–137, 1992.

[BDKR05] Tuğkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating the entropy. *SIAM Journal on Computing*, 35(1):132–150, 2005.

[BFF$^+$01] Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 442–451, 2001.

[BFR$^+$13] Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren Smith, and Patrick White. Testing closeness of discrete distributions. *Journal of the ACM*, 4, 2013.

[BHR99] Bernd Borchert, Lane Hemaspaandra, and Jörg Rothe. Restrictive acceptance suffices for equivalence problems. In *Proceedings of the 12th International Symposium on Fundamentals of Computation Theory*, pages 124–135, 1999.

[BIL12] Christopher Beck, Russell Impagliazzo, and Shachar Lovett. Large deviation bounds for decision trees and sampling lower bounds for $\mathrm{AC}^0$-circuits. In *Proceedings of the 53rd IEEE Symposium on Foundations of Computer Science*, pages 101–110, 2012.

[BKR04] Tuğkan Batu, Ravi Kumar, and Ronitt Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proceedings of the 36th ACM Symposium on Theory of Computing*, pages 381–390, 2004.

[BMV08] Andrej Bogdanov, Elchanan Mossel, and Salil Vadhan. The complexity of distinguishing Markov random fields. In *Proceedings of the 12th International Workshop on Randomization and Computation*, pages 331–342, 2008.

[BS00] 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.

[CDVV13] Siu On Chan, Ilias Diakonikolas, Gregory Valiant, and Paul Valiant. Optimal algorithms for testing closeness of discrete distributions. *CoRR*, abs/1308.3946, 2013.

[DDS$^+$13] Constantinos Daskalakis, Ilias Diakonikolas, Rocco Servedio, Gregory Valiant, and Paul Valiant. Testing $k$-modal distributions: Optimal algorithms via reductions. In *Proceedings of the 24th ACM-SIAM Symposium on Discrete Algorithms*, pages 1833–1852, 2013.

[DF80] Perci Diaconis and David Freedman. Finite exchangeable sequences. *Annals of Probability*, 8(4):745–764, 1980.

[DGRV11] Zeev Dvir, Dan Gutfreund, Guy Rothblum, and Salil Vadhan. On approximating the entropy of polynomial mappings. In *Proceedings of the 2nd Innovations in Computer Science Conference*, pages 460–475, 2011.

[DW12] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Transactions on Computation Theory*, 4(1), 2012.

[FGHP98] Stephen Fenner, Frederic Green, Steven Homer, and Randall Pruim. Quantum NP is hard for PH. In *Proceedings of the 6th Italian Conference on Theoretical Computer Science*, pages 241–252, 1998.

[GSV99] Oded Goldreich, Amit Sahai, and Salil Vadhan. Can statistical zero knowledge be made non-interactive? or On the relationship of SZK and NISZK. In *Proceedings of the 19th International Cryptology Conference*, pages 467–484, 1999.

[GV11] Oded Goldreich and Salil Vadhan. On the complexity of computational problems regarding distributions. *Studies in Complexity and Cryptography*, pages 390–405, 2011.

[IW97] Russell Impagliazzo and Avi Wigderson. P = BPP if E requires exponential circuits: Derandomizing the XOR Lemma. In *Proceedings of the 29th ACM Symposium on Theory of Computing*, pages 220–229, 1997.

[Kin78] John Kingman. Uses of exchangeability. *Annals of Probability*, 6(2):183–197, 1978.

[KRVZ11] Jesse Kamp, Anup Rao, Salil Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. *Journal of Computer and System Sciences*, 77(1):191–220, 2011.

[LRR13] Reut Levi, Dana Ron, and Ronitt Rubinfeld. Testing properties of collections of distributions. *Theory of Computing*, 9:295–347, 2013.

[LV12] Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. *Computational Complexity*, 21(2):245–266, 2012.

[RRSS09] Sofya Raskhodnikova, Dana Ron, Amir Shpilka, and Adam Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM Journal on Computing*, 39(3):813–842, 2009.

[RS09] Ronitt Rubinfeld and Rocco Servedio. Testing monotone high-dimensional distributions. *Random Structures and Algorithms*, 34(1):24–44, 2009.

[RX10] Ronitt Rubinfeld and Ning Xie. Testing non-uniform $k$-wise independent distributions over product spaces. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming*, pages 565–581, 2010.

[SV03] Amit Sahai and Salil Vadhan. A complete problem for statistical zero knowledge. *Journal of the ACM*, 50(2):196–249, 2003.

[Tar93] Jun Tarui. Probabilistic polynomials, $\mathrm{AC}^0$ functions, and the polynomial-time hierarchy. *Theoretical Computer Science*, 113(1):167–183, 1993.

[TO92] Seinosuke Toda and Mitsunori Ogiwara. Counting classes are at least as hard as the polynomial-time hierarchy. *SIAM Journal on Computing*, 21(2):316–328, 1992.

[TVZ05] Luca Trevisan, Salil Vadhan, and David Zuckerman. Compression of samplable sources. *Computational Complexity*, 14(3):186–227, 2005.

[Val11] Paul Valiant. Testing symmetric properties of distributions. *SIAM Journal on Computing*, 40(6):1927–1968, 2011.

[Vio11] Emanuele Viola. Extractors for circuit sources. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science*, pages 220–229, 2011.

[Vio12] Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012.

[Wag86] Klaus Wagner. The complexity of combinatorial problems with succinct input representation. *Acta Informatica*, 23(3):325–356, 1986.

[Wat12] Thomas Watson. The complexity of estimating min-entropy. Technical Report TR12-070, Electronic Colloquium on Computational Complexity, 2012.