The Complexity of Deciding Statistical Properties of Samplable Distributions*

Thomas Watson

University of Toronto, Toronto, Canada
thomasw@cs.toronto.edu

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 $C_=P$-complete, by showing that the following promise problem (which is a restriction of all the above problems) is $C_=P$-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 $C_=P$, which we call $BC_=P$, and we investigate its structural properties.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases Complexity, statistical properties, samplable distributions

Digital Object Identifier 10.4230/LIPIcs.STACS.2014.663

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 [4, 3, 5, 2, 14, 18, 7].

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 set.) For arbitrary circuit samplers, the

* Supported by funding from NSERC.

31st Symposium on Theoretical Aspects of Computer Science (STACS'14).
Editors: Ernst W. Mayr and Natacha Portier; pp. 663–674

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany


papers [15, 10, 11, 22] contain completeness results for various approximation problems concerning statistical distance, Shannon entropy, and min-entropy. See [12] 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 [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. 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 [21] 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 $\mathrm{PH} \subseteq \mathrm{BP} \cdot \mathrm{C_{=}P}$ [17] and even $\mathrm{PH} \subseteq \mathrm{ZP} \cdot \mathrm{C_{=}P}$ [16]. It is no harder than "threshold counting", since $\mathrm{C_{=}P} \subseteq \mathrm{PP}$, but neither is it substantially easier, since $\mathrm{PP} \subseteq \mathrm{NP}^{\mathrm{C_{=}P}}$. It was shown in [9] that $\mathrm{C_{=}P} = \mathrm{coNQP}$.

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, such as parallel time, that are combinatorially simple enough to reason about and obtain unconditional results. 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 [20, 13, 19, 6]. 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 [20, 8, 19, 22] as well as a large collection of papers investigating the possibility of implementing pseudorandom generators locally. (See [8] 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.)

We prove that our $\mathrm{C_{=}P}$-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 or far from satisfying a property. It was shown in [10] that for the property of being uniform, the problem is NISZK-complete. It was shown in [15] that the problem of deciding whether a pair of samplable distributions are statistically close or far is complete for 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 $\mathrm{C}_{=}\mathrm{P}$, which we call $\mathrm{BC}_{=}\mathrm{P}$ 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 $i \in \{1,\ldots,n\}$) 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 $\leq s$ 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 for exact problems in Section 2.1 and prove them in Section 3 and the full version. We state our algorithmic results for exact problems in Section 2.2 and prove them in Section 4 and the full version. We state our completeness results for approximate problems in Section 2.3 and prove them in Section 5 and the full version. In the full version, we also study a new complexity class, $\mathrm{BC}_{=}\mathrm{P}$.

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 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 ≥ 2 coordinates and every pair of coordinates is neither uncorrelated nor identically distributed.

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

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

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 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, $S \in \textsc{Uniform} \iff S(U)$ 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 $\mathrm{C}_{=}\mathrm{P}$. In fact, the $\mathrm{C}_{=}\mathrm{P}$-hardness of each of the individual languages in Fact 1 is fairly simple to prove, but the $\mathrm{C}_{=}\mathrm{P}$-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 $\mathrm{C}_{=}\mathrm{P}$-hardness of all the languages in Fact 1.

To prove the $\mathrm{C}_{=}\mathrm{P}$-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.

$\mathrm{prC}_{=}\mathrm{P}$ 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 $\mathrm{prC}_{=}\mathrm{P}$.

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

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

$\mathrm{C}_{=}\mathrm{P}$ is defined as the class of languages in $\mathrm{prC}_{=}\mathrm{P}$.

▶ **Theorem 6.** Panoptic-Stats$^{\kappa,\nu,\sigma}$ *is* $\mathrm{prC}_{=}\mathrm{P}$*-hard for every polite* κ, ν, σ *with* $\kappa\nu \leq o(\sigma)$.

▶ **Theorem 7.** Panoptic-Stats$^{\kappa,\nu,\sigma}$ *is* $\mathrm{prC}_{=}\mathrm{P}$*-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 8.** *All the languages listed in Fact 1 are in* $\mathrm{C}_{=}\mathrm{P}$.

Consequently, all the languages listed in Fact 1 are $\mathrm{C}_{=}\mathrm{P}$-complete, even when restricted to (κ, ν, σ)-samplers (like in Definition 4) 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).

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 simple to check whether a given sampler satisfies the characterization.)

▶ **Theorem 9.** *There exists a linear-time algorithm for deciding whether the joint distribution of a given 2-local $(1, n)$-sampler is fully independent.*

▶ **Theorem 10.** *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 11.** *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.*

2.3 Approximate Completeness Results

The statistical distance between two distributions $D^{(1)}, D^{(2)}$ over the same set is defined as $\left\|D^{(1)} - D^{(2)}\right\| = \max_{\text{events } E} \left|\Pr[D^{(1)} \in E] - \Pr[D^{(2)} \in E]\right|$. 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 [15].

▶ **Definition 12.** For functions $0 \leq c(k, n, s) < f(k, n, s) \leq 1$, $\textsc{Fully-Independent}^{c,f}$ is the following promise problem.[1]

$$\textsc{Fully-Independent}^{c,f}_{\text{YES}} = \left\{S \;:\; \begin{array}{l} S \text{ is a } (k, n, s)\text{-sampler and } S(U) \text{ is } c(k, n, s)\text{-close} \\ \text{to some fully independent distribution over } (\{0,1\}^k)^n \end{array}\right\}$$

$$\textsc{Fully-Independent}^{c,f}_{\text{NO}} = \left\{S \;:\; \begin{array}{l} S \text{ is a } (k, n, s)\text{-sampler and } S(U) \text{ is } f(k, n, s)\text{-far} \\ \text{from every fully independent distribution over } (\{0,1\}^k)^n \end{array}\right\}$$

$\textsc{Exchangeable}^{c,f}$ is defined in an analogous way.

▶ **Theorem 13.** $\textsc{Fully-Independent}^{c,f}$ *is* prSZK*-hard for all constants* $0 < c < f < \frac{1}{4}$.

▶ **Theorem 14.** $\textsc{Fully-Independent}^{c,f} \in$ prSZK *where* $c = c'/(n+1)$*, for all constants* $0 < c' < f^2 < 1$.

▶ **Theorem 15.** $\textsc{Exchangeable}^{c,f}$ *is* prSZK*-hard for all constants* $0 < c < f < \frac{1}{2}$.

▶ **Theorem 16.** $\textsc{Exchangeable}^{c,f} \in$ prSZK *for all constants* $0 < 2c < f^2 < 1$.

Consequently for example $\textsc{Fully-Independent}^{0.05/(n+1),\ 0.24}$ and $\textsc{Exchangeable}^{0.12,\ 0.49}$ are prSZK-complete.

[1] The superscripts have a different meaning than the superscripts in Definition 4.

3 Proofs of Exact Completeness Results

3.1 The Key Lemma

The following is the key lemma in the proof of Theorem 6. It can be interpreted qualitatively as a certain type of amplification.

▶ **Lemma 17.** *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, ..., a_n uniformly at random
if there exists an ℓ < n such that a_ℓ = 0 then
    let ℓ* be the least such ℓ
    output a_1, ..., a_{ℓ*}, b, a_{ℓ*+2}, ..., a_n
else output a_1, ..., 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 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

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

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

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

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

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

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

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

It follows that

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

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

▶ **Lemma 18.** *Lemma 17 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 17, 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

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

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 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. ◀

3.2 prC$_=$P-Hardness

▶ **Corollary 19.** *Lemmas 17 and 18 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 17) or $O(kn + n^2 + s)$ (for Lemma 18).*

See the full version for the straightforward proof of Corollary 19.

Proof of Theorem 6. We reduce Uniform-Bit to Panoptic-Stats$^{\kappa,\nu,\sigma}$. Let c be the constant factor in the big O in Corollary 19. Given a $(1, 1, s)$-sampler S, we first find the smallest m such that $c \cdot \left(\kappa(m)\nu(m) + s\right) \leq \sigma(m)$. Such an m exists and is $O(s)$ because

$\kappa\nu \le o(\sigma)$ and $\sigma(m) \ge m$ for all m. Then we run the algorithm from Corollary 19 (based on Lemma 17) with $k = \kappa(m)$ and $n = \nu(m)$ to get T of size at most $c \cdot \big(\kappa(m)\nu(m) + s\big) \le \sigma(m)$. Thus the following both hold.

$$S \in \textsc{Uniform-Bit}_{\text{YES}} \implies T \in \textsc{Panoptic-Stats}^{\kappa,\nu,\sigma}_{\text{YES}}$$
$$S \in \textsc{Uniform-Bit}_{\text{NO}} \implies T \in \textsc{Panoptic-Stats}^{\kappa,\nu,\sigma}_{\text{NO}}$$

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 19 runs in time $O(kn + s)$. ◀

The proof of Theorem 7 is similar; see the full version for details.

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

In the proof of Theorem 8 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 the full version.

▶ **Lemma 20.** *If $L \in \mathrm{C_{=}P}$ then both of the following hold.*

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

Proof of Theorem 8. 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 20 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^{th} coordinate of the output of the i^{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 π 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 π. Thus by Lemma 20 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 $L = \big\{(S,i,j) \ : \ S(U)_i \text{ and } S(U)_j \text{ are uncorrelated}\big\}$, then 2-Exists-Uncorrelated

reduces to $\vee L$ by mapping a (k,n)-sampler S to $\big((S,1,2),(S,1,3),(S,1,4),\ldots,(S,n-1,n)\big)$. Thus by Lemma 20 it suffices to show that $L \in \mathrm{C}_{=}\mathrm{P}$. 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 $\Pr[S_1(U)=1] = \frac{1}{2^{2k}} \cdot \mathrm{E}\big(S(U)_i \cdot S(U)_j\big)$, and S_2 runs S twice (independently) yielding some $y^{(1)}$ and $y^{(2)}$ and accepts with probability $\frac{1}{2^{2k}} \cdot y_i^{(1)} \cdot y_j^{(2)}$ so that $\Pr[S_2(U)=1] = \frac{1}{2^{2k}} \cdot \mathrm{E}(S(U)_i) \cdot \mathrm{E}(S(U)_j)$. ◀

4 Proofs of Exact Algorithmic Results

We prove Theorem 9 in Section 4.1. The proof of Theorem 10 (on exchangeability of distributions with 2-local samplers) is much more interesting and less elementary, but due to space constraints we must defer it to the full version (where we also prove Theorem 11).

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 Full Independence for 2-Local Samplers

We prove Theorem 9. 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 $\frac{1}{2^\ell}$ (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 $\frac{1}{2^{\ell+1}}$ (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 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 $\frac{1}{2^\ell}$ (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 $\frac{1}{2^{\ell+1}}$ (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.

5 Proofs of Approximate Completeness Results

Due to space constraints, we defer most of this section to the full version. Here, we just give the argument for Theorem 16, which uses the following lemma.

▶ **Lemma 21.** *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 21. 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 $\{w_i \ : \ i \in \{1, \ldots, n\}\}$ 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 we have

$$\begin{aligned} \|D - D^*\| &= \tfrac{1}{2} \cdot \textstyle\sum_{w \in (\{0,1\}^k)^n} \big|\Pr[D = w] - \Pr[D^* = w]\big| \\ &= \tfrac{1}{2} \cdot \textstyle\sum_{\text{multisets } W \subseteq \{0,1\}^k \text{ of size } n} \big(d_W^{*+} + d_W^{*-}\big) \end{aligned} \tag{1}$$

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

$$\|D - D'\| = \tfrac{1}{2} \cdot \textstyle\sum_{\text{multisets } W \subseteq \{0,1\}^k \text{ of size } n} \big(d_W'^{+} + d_W'^{-}\big). \tag{2}$$

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

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

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

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

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 Equations (1) and (2). ◀

We mention that the constant factor of 2 in Lemma 21 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 $\|D - D^*\| = \frac{1}{n}$. However, D' has probability $\frac{1}{n^2}$ on each string of Hamming weight 1, and probability $\frac{n-1}{n^2}$ on each string of Hamming weight $n - 1$, and thus $\|D - D'\| = 2(1 - \frac{1}{n}) \cdot \frac{1}{n} = 2(1 - \frac{1}{n}) \cdot \|D - D^*\|$.

To prove Theorem 16, we reduce $\textsc{Exchangeable}^{c,f}$ to the promise problem of deciding whether the distributions of two given samplers are $2c$-close or f-far in statistical distance (from each other), which Sahai and Vadhan [15] proved is in prSZK for all constants $0 < 2c < f^2 < 1$. Given a (k, n)-sampler S with distribution $D = S(U)$, the reduction outputs S and another (k, n)-sampler S' that samples from D' (as in the statement of Lemma 21) by running S then permuting the coordinates uniformly at random. (There is a minor technical issue arising from $n!$ not being a power of 2, but this is not problematic.) If D is c-close to some exchangeable distribution then D, D' are $2c$-close. If D is f-far from every exchangeable distribution then D, D' are f-far since D' is exchangeable.

References

1 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.

2 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.

3 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.

4 Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren Smith, and Patrick White. Testing closeness of discrete distributions. *Journal of the ACM*, 4, 2013.

5 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.

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

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

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

9 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.

10 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.

11 Oded Goldreich and Salil Vadhan. Comparing entropies in statistical zero-knowledge with applications to the structure of SZK. In *Proceedings of the 14th IEEE Conference on Computational Complexity*, page 54–73, 1999.

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

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

14 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.

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

16 Jun Tarui. Probabilistic polynomials, AC^0 functions, and the polynomial-time hierarchy. *Theoretical Computer Science*, 113(1):167–183, 1993.

17 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.

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

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

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

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

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