Electronic Colloquium on Computational Complexity, Report No. 4 (2011)

On the complexity of computational problems regarding distributions (a survey)*

Oded Goldreich†
Department of Computer Science
Weizmann Institute of Science
Rehovot, Israel.
oded@wisdom.weizmann.ac.il

Salil Vadhan‡
Division of Engineering and Applied Sciences
Harvard University
Cambridge, MA, USA.
salil@eecs.harvard.edu.

December 30, 2010

Abstract

We consider two basic computational problems regarding discrete probability distributions: (1) approximating the statistical difference (aka variation distance) between two given distributions, and (2) approximating the entropy of a given distribution. Both problems are considered in two different settings. In the first setting the approximation algorithm is only given samples from the distributions in question, whereas in the second setting the algorithm is given the "code" of a sampling device (for the distributions in question).

We survey the know results regarding both settings, noting that they are fundamentally different: The first setting is concerned with the number of samples required for determining the quantity in question, and is thus essentially information theoretic. In the second setting the quantities in question are determined by the input, and the question is merely one of computational complexity. The focus of this survey is actually on the latter setting. In particular, the survey includes proof sketches of three central results regarding the latter setting, where one of these proofs has only appeared before in the second author's PhD Thesis.

Keywords: Approximation, Reductions, Entropy, Statistical Difference, Variation Distance, Sampleable Distributions, Zero-Knowledge, and Promise Problems.

*This survey was first drafted in 2003.
†Written while being at the Radcliffe Institute for Advanced Study of Harvard University.
‡Written while being at the Radcliffe Institute for Advanced Study of Harvard University.

ISSN 1433-8092

1 Introduction

We consider two basic computational problems regarding discrete probability distributions:

1. Computing (or rather approximating) the statistical difference (aka variation distance) between two given distributions.

2. Computing (or rather approximating) the entropy of a given distribution.

The foregoing informal phrases avoid the question of representation; that is, how are the distributions given to the algorithms. Both computational problems are quite trivial in the case that the distributions are explicitly given to the algorithm (i.e., by a list of all elements in the support of the distribution coupled with the probability mass assigned to them). Very good additive approximations can be obtained also in the case that the algorithm is given sufficiently many samples (drawn independently) from the distribution, where "sufficiently many" means linear in the size of the distribution's support. For example, given $N/\mathrm{poly}(\epsilon)$ samples from a distribution that has support size (at most) N, one can estimate the distribution's entropy up-to an additive deviation of ϵ (w.v.h.p.). The same number of samples suffices for approximating the statistical distance between two such distributions (again, up to an additive deviation of ϵ, w.v.h.p.).

The question is whether such approximations (or even weaker ones) can be obtained based on significantly less samples. At the very least, we are interested in algorithms that take $o(N)$ samples (i.e., a "sub-linear" (in the support size) number of samples). In Section 3, we survey what is known regarding this question. The bottom-line is that weak approximations of both quantities can be obtained using N^e samples, for some $e < 1$, but nothing significant can be achieved with $N^{o(1)}$ samples.

We note that the foregoing question is essentially an information-theoretical one; that is, the question refers to the number of samples required to make some estimations regarding the distribution(s). In contrast, in Section 4, we consider a purely computational-complexity problem: We consider algorithms that are given the "code" of a sampling device (for the distributions in question). We stress that such a device fully determines the distribution (from an information-theoretic point of view), and the issue is what quantities can be *efficiently* computed based on this description of the distribution. Note that the algorithm may, of course, use the sampling device in order to generate samples. However, the algorithm is not confined to this usage of the sampling device and may try to analyze the device in other ways (e.g., try to "reverse-engineer" it).

To be concrete, the sampling device is represented by a circuit $C : \{0,1\}^m \to \{0,1\}^n$, which can be used to generate samples by feeding it with a uniformly selected m-bit long string. Alternatively, one may say that C is an implicit representation of a distribution over $\{0,1\}^n$, obtained by feeding C with a uniformly selected m-bit long string. Typically, the circuit's size is polynomial in n, whereas the distribution defined by it can have support size 2^n. Thus, when we consider the aforementioned computational problems in terms of the circuit size, polynomial-time algorithms correspond to algorithms that run in time that may be poly-logarithmic in the size of the support. We stress that, in this model, the algorithm has full information regarding the distribution in question, but it does not have enough time to use this information in a straightforward way (i.e., feed the circuit with all possible inputs). The question is whether the algorithm can obtain approximations to the aforementioned quantities within time that is polynomial in the size of the circuit. In Section 4, we survey what is known regarding this question. The bottom-line is that the complexity of approximating each of the foregoing computational problems is complete (under polynomial-time reductions) for the complexity class $\mathcal{SZK} \subseteq \mathcal{AM} \cap \mathrm{co}\mathcal{AM}$, which is conjectured to extend beyond $\mathcal{BPP}$ (i.e., probabilistic polynomial-time). In particular, under the widely believed conjecture that

the Discrete Logarithm Problem is intractable, it follows that the approximation versions of each of the foregoing computational problems are intractable. It is also known that the two types of computational problems are actually computationally equivalent; that is, each is efficiently reducible to the other.

Organization: In Section 3 we briefly survey the known results regarding sampling-based algorithms (i.e., algorithms that only get samples from the distributions in question). In Section 4 we survey the known results regarding the second setting; that is, we consider algorithms that are given as input a full description of a sampling device for the distributions in question. In Section 5 we present the main ideas underlying the proofs of the three theorems stated in Section 4. One of these proofs has only appeared before in the second author's PhD Thesis [22]. Sections 4 and 5 are actually the main part of this survey.

2 Preliminaries

Traditionally, (discrete) probability distributions are represented by the list of probabilities assigned to the various elements in their range (or potential support). That is, a distribution is presented by a sequence $(p_1, ..., p_N)$ of non-negative numbers (which sum-up to one) such that p_i represents the probability mass that is assigned to the ith element, denoted e_i. Without loss of generality, we may assume that $\{e_i : i = 1, ..., N\} = [N] \stackrel{\text{def}}{=} \{1, ..., N\}$.

In this survey, we prefer to represent probability distributions by corresponding random variables that represent an element selected according to the distribution in question. That is, for a sequence $(p_1, ..., p_N)$ as above, we consider a random variable $X \in [N]$ such that $p_i = \Pr[X = e_i]$, and identify the random variable X with the probability distribution that assigns to e_i the probability mass $\Pr[X = e_i]$.

The statistical difference (or variation distance) between the distributions (or the random variables) X and Y is defined as

$$\Delta(X, Y) \stackrel{\text{def}}{=} \frac{1}{2} \cdot \sum_e |\Pr[X = e] - \Pr[Y = e]| \;=\; \max_S \{\Pr[X \in S] - \Pr[Y \in S]\} \tag{1}$$

We say that X and Y are δ-close if $\Delta(X, Y) \leq \delta$ and that they are δ-far if $\Delta(X, Y) \geq \delta$. Note that X and Y are identical if and only if they are 0-close, and are disjoint (or have disjoint support) if and only if they are 1-far.

The entropy of a distribution (or random variables) X is defined as

$$\mathrm{H}(X) \stackrel{\text{def}}{=} \sum_e \Pr[X = e] \cdot \log_2(1/\Pr[X = e]) \,. \tag{2}$$

The entropy of a distribution is always non-negative and is zero if and only if the distribution is concentrated on a single element. In general, a distribution that has support size N has entropy at most $\log_2 N$.

3 Sampling-based algorithms

In this section we consider algorithms that approximate quantities related to distributions solely on the basis of samples of the relevant distributions. We refer to such algorithms as sampling-based algorithms, and consider such algorithms for approximating the distance between pairs of

distributions and approximating the entropy of a distribution. We denote by N an upper bound on the size of the support of these distributions, and focus on algorithms that obtain $o(N)$ samples.

We review the known results regarding the relationship between the number of samples and the quality of the approximation. In other words, we consider the sample complexity of these approximation problems.

3.1 Approximating the distance between distributions

The study of sampling-based algorithms for approximating the statistical distance between distributions was initiated by Batu *et. al.* [6]. They show that $\Theta(N^{1/2})$ samples are necessary and sufficient in order to distinguish a pair of identical distributions from a pair of disjoint distributions (i.e., to distinguish the case that the two distributions are 0-close from the case that they are 1-far), where N is an upper bound on the support of the distribution. Regarding the more general problem of distinguishing pairs of identical distributions from pairs of distributions that are δ-far, Batu *et. al.* [6] showed that $\widetilde{O}(N^{2/3}\delta^{-4})$ samples suffice, and claimed that $\Omega(N^{2/3})$ samples are necessary. The latter claim was proved by P. Valiant [25]. Regarding the even more general problem of approximating the statistical distance between distributions, it was shown by P. Valiant [25] that $N^{1-o(1)}$ samples are required. That is, for every fixed $0 < \delta_1 < \delta_2 < 1$, it is the case that $N^{1-o(1)}$ samples are required in order to distinguish distribution-pairs that are δ_1-close from distribution-pairs that are δ_2-far apart.

Our conclusion is that in order to obtain any meaningful information regarding the distance between two distributions (in this model), one must obtain $\Omega(N^{1/2})$ samples. Furtherthmore, while $O(N^{2/3})$ samples suffice for distinguishing identical distribution-pairs from distribution-pairs that are far apart (say 0.1-far), in the general case $N^{1-o(1)}$ samples are required in order to approximate (up to any constant additive term) the statistical distance between two distributions (of support size N).

3.2 Approximating the entropy of a distribution

Batu *et. al.* [4] considered the problem of approximating the entropy of a distribution based on samples from it; that is, they considered sampling-based algorithms for this task. They presented an algorithm that, for any $\gamma > 1$, using $\widetilde{O}(N^{1/\gamma})$ samples of a distribution that has entropy $\Omega(\gamma)$ provides a γ-factor approximation of its entropy. We comment that some lower-bound on the entropy is necessary for obtaining any approximation-factor based on samples.[1] On the other hand, also in the case that the entropy is lower-bounded (as in Footnote 1 or even more), a constant factor approximation of the entropy requires $N^{\Omega(1)}$ samples (i.e., a γ-factor approximation requires $\Omega(N^{(1/\gamma)-o(1)})$ samples; see [25]).

Our conclusion is that, except in pathological cases (of distributions having very small entropy), the sample complexity of obtaining a γ-factor approximation of the entropy of a distribution is $N^{(1/\gamma)\pm o(1)}$, where N is an upper bound on the support of the distribution.

Additive error approximation. The foregoing discussion refers to multiplicative error approximation. Recent work by G. Valiant and P. Valiant [23, 24] refers to additive error approximations and shows that $\Theta(n/\log n)$ samples are necessary and sufficient in such a case.

[1] Consider, for example, the family of distributions (parameterized by $\epsilon > 0$) having support size 2, assigning probability ϵ to one element and $1 - \epsilon$ to the other.

3.3 Additional comments

A general framework for analyzing the sample complexity of various computational problems regarding distributions was recently provided by P. Valiant [25]. Indeed, some of the aforementioned lower-bounds are derived using this framework. Furthermore, this framework may be applied to other natural measures of distance between distributions.

Some of the aforementioned results can be cast naturally within the formalism of property testing (cf. [20, 12, 9]). For example, one may consider the property of two distributions being identical, and the task of accepting pairs having the property and rejecting pairs that are far from having the property according to a natural distance measure (cf. [9]).

Related work. Batu *et. al.* [5] have considered the task of approximating the distance between a fixed distribution and a second distribution for which one only obtains samples.[2] They present an algorithm that, for a parameter δ, determines whether the two distributions are $\mu(N) \cdot \delta^3$-close or δ-far based on $\widetilde{O}(N^{1/2}\delta^{-O(1)})$ samples, where $\mu(N) = \widetilde{O}(1/\sqrt{N})$. This matches a lower bound of $\Omega(\sqrt{N})$ samples requires to distinguish the case that the distribution is uniform over $[N]$ from the case that it is (say) 0.1-far from being uniform. Batu *et. al.* [4] considered the problem of approximating the entropy of a distribution also in a model in which the algorithm has access to an "evaluation oracle" instead or in addition to the samples, where the evaluation oracle is defined to answer the query x with the probability mass assigned to x.

4 Algorithms that are given a sampling device

In this section we consider algorithms that are given a succinct description of the distributions in question. That is, the algorithm is given a "sampling device" (in the form of a circuit) and is supposed to approximate a quantity that refers to the distribution defined by this sampling device. A sampling device is actually an algorithm, and the distribution defined by it is the output distribution of the device when fed with a random input of adequate length. For concreteness, for a feasibility parameter n, we consider $\mathrm{poly}(n)$-size circuits that map $\mathrm{poly}(n)$-bit long inputs to n-bit long outputs. Note that such circuits define a distribution over $\{0,1\}^n$, which may contain $N = 2^n$ elements. In other words, a distribution over $\{0,1\}^n$ is represented by a corresponding ($\mathrm{poly}(n)$-size) sampling device (or circuit), which typically means that we use a succinct representation of the distribution.

We consider algorithms that are given such a representation (i.e., a circuit) as input, and need to approximate some quantities of the represented distribution. Indeed, one thing that such an algorithm can do is evaluate the circuit on inputs of its choice, and in particular on uniformly selected inputs. Thus, the algorithm can certainly produce samples of the distribution, where these samples are indeed of the type used in Section 3. However, the algorithm is not confined to operating in that way, and it may try to "reverse engineer" the circuit in order to learn more about the distribution (than by merely observing random samples generated according to the distribution). Needless to say, we don't really believe that "reverse engineering" can help to answer the computational problems considered here, still we cannot rule out this possibility.

We stress that unlike in Section 3, the algorithm gets full information of the distribution. That is, from an information theoretic point of view, the sampling device (or circuit) determines the distribution, and thus determines its entropy and its distance from another distribution. The

[2] Alternatively, the first distribution may be given explicitly (as input to the algorithm), which in this case has running time linear in N.

question is how much time is required in order to compute these quantities from the information that fully-determines them. In the rest of this section we associate the sampling circuits with the distributions generated by them. That is, we associate the circuit C with the distribution it outputs when fed with a uniformly selected input.

We study the complexity of approximation problems by defining corresponding promise problems (cf. [7]), where the latter are pairs of disjoint sets (cf. [10]). A promise problem (A, B) consists of distinguishing between inputs in A and inputs in B, where inputs out of $A \cup B$ are ignored (or one is "promised" that the input is in $A \cup B$).

We briefly recall the standard definitions of reductions, when applied to promise problems. The promise problem (A_1, B_1) is Karp-reducible to (A_2, B_2) if there exists a polynomial-time computible function f such that if $x \in A_1$ (resp., $x \in B_1$) then $f(x) \in A_2$ (resp., $f(x) \in B_2$). More generally, (A_1, B_1) is Cook-reducible (or just reducible) to (A_2, B_2) if there exists a polynomial-time oracle machine M that on input $x \in A_1$ (resp., $x \in B_1$) and oracle access to (A_2, B_2), outputs 1 (resp., 0), where query q to the oracle (A_2, B_2) is answered arbitrarily in case $q \notin A_2 \cup B_2$. Two problems are said to be computationally equivalent (resp., computationally equivalent under Karp-reductions) if each is Cook-reducible (resp., Karp-reducible) to the other.

4.1 Approximating the distance between distributions

We consider promise problems that take as input a pair of circuits and refer to the statistical difference between the two corresponding distributions (generated by the two circuits). For (threshold) functions $c, f : \mathsf{N} \to [0,1]$, where $c \leq f$, the promise problem $\mathtt{GapSD}^{c,f} = (\mathtt{Close}^c, \mathtt{Far}^f)$ is defined such that $(C_1, C_2) \in \mathtt{Close}^c$ if $\Delta(C_1, C_2) \leq c(|C_1| + |C_2|)$ and $(C_1, C_2) \in \mathtt{Far}^f$ if $\Delta(C_1, C_2) > f(|C_1| + |C_2|)$. In particular, we focus on promise problem $\mathtt{GapSD} \stackrel{\text{def}}{=} \mathtt{GapSD}^{\frac{1}{3},\frac{2}{3}}$. Interestingly, the complexity of this gap problem, which captures a moderately good approximation requirement, is computationally equivalent to a very crude approximation requirement. That is, the former problem is Karp-reducible to the latter:

Theorem 1 ([21], see proof sketch in Section 5.1:) *There exists a Karp-reduction of* $\mathtt{GapSD}^{\frac{1}{3},\frac{2}{3}}$ *to* $\mathtt{GapSD}^{\epsilon,1-\epsilon}$*, where* $\epsilon(n) = 2^{-n}$*. More generally, for every polynomial-time computible* $c, f : \mathsf{N} \to [0,1]$ *such that* $c(n) < f(n)^2 - (1/\mathrm{poly}(n))$ *it holds that* $\mathtt{GapSD}^{c,f}$ *is Karp-reducible to* $\mathtt{GapSD}^{\epsilon,1-\epsilon}$*.*

Using a trivial reduction in the other direction, we conclude that *for every* $c, f : \mathsf{N} \to [0,1]$ *such that* $c(n) \geq 2^{-n}$*,* $c(n) < f(n)^2 - (1/\mathrm{poly}(n))$ *and* $f(n) \geq 1 - 2^{-n}$*, the problems* $\mathtt{GapSD}^{c,f}$ *and* $\mathtt{GapSD} = \mathtt{GapSD}^{\frac{1}{3},\frac{2}{3}}$ *are computationally equivalent* (under Karp reductions). This equivalence is useful in determining the complexity of GapSD (as well as all these $\mathtt{GapSD}^{c,f}$'s). Sahai and Vadhan [21] showed that any promise problem having a statistical zero-knowledge proof system is Karp-reducible to $\mathtt{GapSD}^{\frac{1}{2p^2},\frac{1}{p}}$, for some polynomial p, and that $\mathtt{GapSD}^{\epsilon,1-\epsilon}$ (where $\epsilon(n) = 2^{-n}$) has a statistical zero-knowledge proof system. Denoting the class of promise problem having statistical zero-knowledge proof systems by $\mathcal{SZK}$, we have:

Theorem 2 [21]: *The promise problem* GapSD *is* $\mathcal{SZK}$*-complete* (under Karp-reductions)*.*

Recall that $\mathcal{SZK}$ contains some promise problems (e.g., one equivalent to Discrete Logarithm Problems) that are widely believed not to be in $\mathcal{BPP}$ (cf. [13]). On the other hand, $\mathcal{SZK} \subseteq \mathcal{AM} \cap \mathrm{co}\mathcal{AM}$ (cf. [11, 1]), which in turn is quite low in the Polynomial-Time Hierarchy.

We comment that $\mathtt{GapSD} = (\mathtt{Close}, \mathtt{Far})$ *is Karp-reducible to its complement* $(\mathtt{Far}, \mathtt{Close})$ [21]; that is, there is a Karp-reduction that maps pairs (C_1, C_2) to pairs (C_1', C_2') such that if $\Delta(C_1, C_2) \leq 1/3$ then $\Delta(C_1', C_2') > 2/3$ whereas if $\Delta(C_1, C_2) > 2/3$ then $\Delta(C_1', C_2') \leq 1/3$.

4.2 Approximating the entropy of a distribution

We consider two computational problems related to approximating the entropy of a distribution. The first problem is captured by promise problems that take as input a circuit and a value and refers to the relation between the entropy of (the distribution generated by) the circuit and the given value. For a (slackness) function $s : \mathsf{N} \to \mathsf{R}$, where $s > 0$, the promise problem $\mathtt{GapEnt}^s = (\mathtt{Smaller}^s, \mathtt{Larger})$ is defined such that $(C, v) \in \mathtt{Smaller}^s$ if $\mathrm{H}(C) \leq v - s(|C|)$ and $(C, v) \in \mathtt{Larger}$ if $\mathrm{H}(C) \geq v$. In particular, we focus on promise problem $\mathtt{GapEnt} \stackrel{\text{def}}{=} \mathtt{GapEnt}^1$ (which refers to approximating the entropy up to an additive error of 1). It is easy to see that, for every polynomial p and for every $\epsilon > 0$ and $\ell(n) = n^{1-\epsilon}(n)$, the problems $\mathtt{GapEnt}^{1/p}$, $\mathtt{GapEnt}^1$ and $\mathtt{GapEnt}^\ell$ are computational equivalent (under Karp reductions).[3]

We also consider promise problems that take as input a pair of circuits and refer to the relation between the entropies of the corresponding distributions (generated by the two circuits). For a (slackness) function $s : \mathsf{N} \to \mathsf{R}$, where $s > 0$, the promise problem $\mathtt{GapCmprEnt}^s = (\mathtt{Smaller}^s, \mathtt{Larger}^s)$ is defined such that $(C_1, C_2) \in \mathtt{Smaller}^s$ if $\mathrm{H}(C_1) \leq \mathrm{H}(C_2) - s(|C_1| + |C_2|)$ and $(C_1, C_2) \in \mathtt{Larger}^s$ if $\mathrm{H}(C_1) \geq \mathrm{H}(C_2) + s(|C_1| + |C_2|)$. In particular, we focus on promise problem $\mathtt{GapCmprEnt} \stackrel{\text{def}}{=} \mathtt{GapCmprEnt}^1$, and note that it is computationally equivalent (under Karp reductions) to $\mathtt{GapCmprEnt}^{1/p}$ and $\mathtt{GapCmprEnt}^\ell$ (where p and ℓ are as above). Two easy observations follow:

Observation 1: *The problems* GapEnt *and* GapCmprEnt *are computationally equivalent* (under Cook reductions). Specifically, GapEnt is Karp-reducible to GapCmprEnt, whereas GapCmprEnt is Cook-reducible to GapEnt.

For example, one may use a Karp-reduction that maps an instance (C, v) of GapEnt to the intance $(C, C_{v-0.5})$ of $\mathtt{GapCmprEnt}^{1/3}$ such that $C_{v-0.5}$ is a standard circuit that generates some distribution of entropy (approximately) $v - 0.5$. For the other direction, consider an oracle machine that decides intances of GapCmprEnt by using queries to $\mathtt{GapEnt}^{1/3}$ in order to determine the entropy of each of the two input distributions (up to an additive error of 1/3).

Observation 2: *The problem* $\mathtt{GapCmprEnt} = (\mathtt{Smaller}, \mathtt{Larger})$ *is Karp-reducible to its complement* $(\mathtt{Larger}, \mathtt{Smaller})$; e.g., by the reduction that maps (C_1, C_2) to (C_2, C_1).

It is not know whether or not GapCmprEnt is Karp-reducible to GapEnt and whether or not GapEnt is Karp-reducible to its complement. In fact, both questions are equivalent (cf. [16]), and we conjecture that the answer (to both of them) is negative. It turns out that all these computational problems (regarding entropy) are computationally equivalent to the computational problems regarding statistical distance:

Theorem 3 ([17], see proof sketch in Section 5.3:) *The promise problems* GapCmprEnt *and* GapSD *are computationally equivalent under Karp reductions.*

Combining Theorem 3 and Observation 2, it follows that $\mathtt{GapSD} = (\mathtt{Close}, \mathtt{Far})$ *is Karp-reducible to its complement* $(\mathtt{Far}, \mathtt{Close})$. We comment that this result (which was already stated at the end of Subsection 4.1) was originally proved in [21] without using the equivalence of GapSD and GapCmprEnt (i.e., without using Theorem 3).

[3] The tighter (additive) approximation is reduced to the looser one by combining sufficiently many copies of the circuit.

4.3 Additional comments

We comment that the promise problems GapSD, GapEnt and GapCmprEnt were originally introduced as tools in the study of statistical zero-knowledge.[4] Consequently, the original presentations (cf. [21, 17, 16]) focus on the derivation and presentation of results regarding statistical zero-knowledge, and the relation between the promise problems themselves is sometimes only implicit (and is typically not at the main focus). In fact, redeeming this state of affairs has been our initial motivation for writing the current survey.

The bottom-line of the foregoing results is that many of the approximation versions of the two problems (i.e., approximating the distance between distributions and approximating their entropy) are computationally-equivalent. The exceptional versions that are not known to be equivalent to the other versions refer to too small gaps (which may yield even harder versions). Whereas in the case of approximating the entropy the definition of "too small gaps" is a natural one, it is somewhat artificial in the case of $\mathtt{GapSD}^{c,f}$ where we require $c < f^2$. An interesting open problem is to determine the complexity of $\mathtt{GapSD}^{c,f}$ in the case that $c > f^2$ (but $c < f$, of course)[5]; that is, is this problem computationally equivalent to GapSD or is it strictly harder?

An alternative perspective on the current section is that it concerns only probability distributions that have a succinct representation, where such a representation is one allowing to efficiently obtain samples from the distribution. Specifically, for a feasibility parameter n, we consider probability distributions over $\{0,1\}^n$. The support of such a distribution may contain 2^n elements, while we consider algorithms operating in poly(n)-time. Thus, such algorithms cannot read an explicit representation of the distribution (in the form of a sequence of length 2^n), and hence the distribution is given to it in a succinct representation. Specifically, we have considered algorithms that are given a sampling device, which is a poly(n)-size circuit that when feed with a random input output a sample that is distributed according to the distribution. We have considered the complexity of estimating various quantities of distributions given by such a succinct representation.

5 Proof sketches for the three theorems

In this section we outline the main ideas used in the proofs of the three theorems stated in Section 4. Theorem 2 is the only one that refers to statistical zero-knowledge and its proof is the only one that assumes any familiarity with zero-knowledge. The other two proofs are based merely on elementary results from probability theory and probabilistic analysis.

As in Section 4, we associate the sampling circuits with the distributions generated by them. That is, we associate the circuit C with the distribution it outputs when fed with a uniformly selected input.

5.1 Proof sketch for Theorem 1

Theorem 1 was proven by Sahai and Vadhan [21], and here we provide an outline of their proof. Recall that the theorem claims a Karp-reduction of $\mathtt{GapSD}^{\frac{1}{3},\frac{2}{3}}$ (or any adequate $\mathtt{GapSD}^{c,f}$) to $\mathtt{GapSD}^{\epsilon,1-\epsilon}$, where $\epsilon(n) = 2^{-n}$. This reduction (called the *Polarization Lemma* in [21]) has the interesting effect of "polarizing the situation": pairs of distributions that are somewhat close (e.g., are at most at distance 1/3 apart) are mapped to pairs of almost identical distributions (i.e., having negligible

[4] For more details regarding statistical zero-knowledge see either [21, 15, 17, 16] or [22].

[5] The above formulation refers to constant c and f. For $c, f : \mathsf{N} \to [0,1]$, we have to require that $c(n) < f(n) - (1/p(n))$ for some polynomial p.

distance between them), whereas pairs of distributions that are somewhat far apart (e.g., at distance at least 2/3) are mapped to pairs of distributions that are very different (e.g., have distance negligiblly close to 1). The "polarizing" reduction is obtained by composing three Karp-reductions, which in turn are of two types. These two types of Karp-reductions (among $\mathtt{GapSD}^{c,f}$ problems) are described next, starting with the simpler one.

The Direct Product reduction: This reduction increases both bounds in the definition of $\mathtt{GapSD}^{c,f}$ (but not in a tight manner). For any (polynomial) t, we reduce $\mathtt{GapSD}^{c,f}$ to $\mathtt{GapSD}^{t\cdot c,1-2\exp(-t\cdot f^2/2)}$ by constructing circuits that generate t samples of each of the corresponding input distributions. That is, we map the circuit pair (C_1,C_2) to (C'_1,C'_2), where $C'_i(r_1,...,r_t) \stackrel{\text{def}}{=} (C_i(r_1),...,C_i(r_t))$. Clearly, the statistical distance between the distributions grows by at most a factor of t. On the other hand, it can be shown that if two distributions are at distance δ then the statistical difference between their t-products is at least $1-2\exp(-t\cdot\delta^2/2)$. (Indeed, it is not true that the statistical difference between the t-products is exactly $t\cdot\delta$, the latter is merely an upper bound on the former.)[6]

The XOR reduction: This reduction decreases both bounds (in a tight manner). For any (polynomial) t, we reduce $\mathtt{GapSD}^{c,f}$ to $\mathtt{GapSD}^{c^t,f^t}$ by mapping the circuit pair (C_0,C_1) to (C'_0,C'_1), where

$$C'_i(b_1,...,b_{t-1},r_1,...,r_{t-1},r_t) \stackrel{\text{def}}{=} \left(C_{b_1}(r_1),...,C_{b_{t-1}}(r_{t-1}),C_{i+\sum_{j=1}^{t-1}b_j \bmod 2}(r_t)\right) .$$

That is, the two output circuits (i.e., the C'_i's) select samples from the two input distributions (respresented by the C_i's), and differ only in the parity of the number samples taken from the (say) first input distribution. Specifically, C'_0 (resp., C'_1) takes an even (resp., odd) number of samples from C_1. It can be shown that if two input distributions are at distance δ then the statistical difference between the constructed (output) distributions is exactly δ^t. (Intuitively, a single sample drawn for one of the two input distributions corresponds to a "weak" encryption of a bit, whereas a sample drawn from one of the output circuits corresponds to encrypting a bit by applying "weak" encryptions to a random sequence of bits that have the desired parity. The "weakness" of the resulting encryption decays exponentially with t; cf. [26].)[7]

We now turn to the actual reduction of $\mathtt{GapSD}^{\frac{1}{3},\frac{2}{3}}$ (or any adequate $\mathtt{GapSD}^{c,f}$) to $\mathtt{GapSD}^{\epsilon,1-\epsilon}$, where $\epsilon(n)=2^{-n}$. This reduction is composed of the following three reductions:

1. A Karp-reduction of $\mathtt{GapSD}^{\frac{1}{3},\frac{2}{3}}$ (or any $\mathtt{GapSD}^{c,f}$ such that $c(n) < f(n)^2 - \frac{1}{\mathrm{poly}(n)}$) to some $\mathtt{GapSD}^{c',f'}$ such that $f'(n) > \sqrt{8n\cdot c'(n)}$.

   Specifically, for an adequate parameter t, we use the XOR reduction and get $c'=c^t$ and $f'=f^t$, which satisfies the desired condition (regarding c' and f') provided that $c<f^2$ (or

[6] The lower bound of $1-2\exp(-t\cdot\delta^2/2)$ can be proved by referring to the second definition in Eq. (1). Specifically, for an adequate set S, it holds that $p \stackrel{\text{def}}{=} \Pr_r[C_1(r)\in S] = \Pr_r[C_2(r)\in S]-\delta$. Thus, C'_1 (resp., C'_2) is expected to have $t\cdot p$ (resp., $t\cdot(p+\delta)$ elements in S. By applying a Chernoff Bound, we note that with probability at least $1-\exp(-t\cdot\delta^2/2)$, the output of C'_1 (resp., C'_2) will have less than $t\cdot(p+\frac{\delta}{2})$ (resp., more than $t\cdot(p+\frac{\delta}{2})$) elements in S. This yields a set S' that demonstrates the claimed lower bound on the statistical difference between C'_1 and C'_2.

[7] Alternatively, consider the following problem. For pairs of random variables, (X_0,X_1) and (Y_0,Y_1), we define a new pair of random variables, (Z_0,Z_1), such that $Z_i=(X_b,Y_{i\oplus b})$, where $b\in\{0,1\}$ is uniformly distributed. Using the first definition in Eq. (1) and expanding the expression for $\Delta(Z_0,Z_1)$, one can show that $\Delta(Z_0,Z_1)=\Delta(X_0,X_1)\cdot\Delta(Y_0,Y_1)$. The general claim (stated above) follows by induction on t.

actually $c(n) < (8n)^{-t/2} \cdot f(n)^2$). In particular, for $c = 1/3$ and $f = 2/3$, we set $t = O(\log n)$ and reduce $\mathtt{GapSD}^{c,f}$ to $\mathtt{GapSD}^{c',f'}$, where $c'(n) \stackrel{\text{def}}{=} c^t = 1/\mathrm{poly}(n)$ and $f'(n) \stackrel{\text{def}}{=} f^t = (f^2/c)^{t/2} \cdot c^{t/2} > \sqrt{8n \cdot c'(n)}$. In general, we set $t = \mathrm{poly}(n)$ such that $(f(n)^2/c(n))^{t/2} \geq 8n$, which is possible because $\frac{f(n)^2}{c(n)} > 1 + \frac{1}{p(n)}$ for some positive polynomial p.

2. A Karp-reduction of a $\mathtt{GapSD}^{c',f'}$ (with c' and f' as obtained in Step 1) to $\mathtt{GapSD}^{c'',f''}$, where $c''(n) = 1/4$ and $f''(n) \geq 1 - 2\exp(-n)$.

   Specifically, for an adequate parameter t (i.e., $t = 1/4c'(n)$), we use the Direct Product reduction and get $c''(n) \stackrel{\text{def}}{=} t \cdot c'(n) = 1/4$ and $f''(n) \stackrel{\text{def}}{=} 1 - 2\exp(-t \cdot f'(n)^2/2)$. Using the hypothesis $f'(n) \geq \sqrt{8n \cdot c'(n)}$, it follows that $f''(n) = 1 - 2\exp(-f'(n)^2/8c'(n)) \geq 1 - 2\exp(-n)$,

3. A Karp-reduction of a $\mathtt{GapSD}^{c'',f''}$ (with c'' and f'' as obtained in Step 2) to $\mathtt{GapSD}^{\epsilon,1-\epsilon}$, where $\epsilon(n) = 2^{-n}$.

   Specifically, we apply the XOR reduction again, but this time with $t = n/2$, and use $(1/4)^t = 2^{-n} = \epsilon(n)$ and $(1 - 2\exp(-n))^t > 1 - 2^{-n} = 1 - \epsilon(n)$.

Combining the above three reductions, we obtain a Karp-reduction of $\mathtt{GapSD}^{\frac{1}{3},\frac{2}{3}}$ (or any $\mathtt{GapSD}^{c,f}$ such that $c(n) < f(n)^2 - \frac{1}{\mathrm{poly}(n)}$) to $\mathtt{GapSD}^{\epsilon,1-\epsilon}$, where $\epsilon(n) = 2^{-n}$.

On the use of the condition $c < f^2$ in the current reduction: Note that in Step 2 we have assumed that $f'(n) \geq \sqrt{8n \cdot c'}$, where (by Step 1) $f' = f^t$ and $c' = c^t$. It follows that we must have $f(n)^t \geq (8n)^{t/2} \cdot (\sqrt{c(n)})^t$, and in particular $f(n)^2 > c(n)$. As discussed in Section 4.3, it is an open problem whether or not there exists an alternative reduction that uses a more relaxed condition (regarding c and f).

5.2 Proof sketch for Theorem 2

Theorem 2 was also proven by Sahai and Vadhan [21], and here we sketch the ideas underlying their proof. The proof consists of two parts: (1) showing that GapSD has a statistical zero-knowledge proof system, and (2) showing that any problem in $\mathcal{SZK}$ is Karp-reducible to GapSD. We try to present the proof ideas while assuming only a superficial familiarity with the notion of statistical zero-knowledge proof systems. A reader that does not feel comfortable with this assumption is invited to skip the current subsection.

The problem GapSD has a statistical zero-knowledge proof system: Using Theorem 1, it suffices to show such a proof system for $\mathtt{GapSD}^{\epsilon,1-\epsilon}$, where $\epsilon(n) = 2^{-n}$. Actually, we present such a proof system for the complement problem (i.e., $(\mathtt{Far}^{1-\epsilon}, \mathtt{Close}^{\epsilon})$), and rely on the (highly non-trivial) fact that GapSD is reducible to its complement.[8] Employing the same idea as in [18, 14], the verifier selects one of the input distributions at random and presents the prover with a random sample generated according to this distribution. The verifier accepts if and only if the prover correctly identifies the distribution from which the sample was taken. Observe that if the input distributions are far apart then the prover can answer correctly with very high probability. On the

[8] As mentioned in Section 4, this fact follows by combining Theorem 3 with Observation 2. An alternative proof of the fact that GapSD is reducible to its complement was given in [21]. (Actually this alternative proof was discovered before Theorem 3.)

other hand, if the input distributions are very close then the prover cannot guess the correct answer with probability significantly larger than 1/2. This establishes that the protocol is an interactive proof (and thus that GapSD is in co$\mathcal{AM}$). It can be shown that this protocol is actually statistical zero-knowledge, intuitively because the verifier learns nothing from the prover's correct answer which is a priori known to to the verifier (in case the two distributions are far apart).

Any problem in $\mathcal{SZK}$ is Karp-reducible to GapSD: We rely on Okamoto's Theorem by which any problem in $\mathcal{SZK}$ has a *public-coin* statistical zero-knowledge proof system. (We comment that an alternative proof of that theorem has subsequently appeared in [17].) We consider an arbitrary (*public-coin*) statistical zero-knowledge proof system. Following Fortnow [11], we observe a discrepency between the behavior of the simulator on YES-instances versus NO-instances:

- In case the input is a YES-instance, the simulator outputs transcripts that are very similar to those in the real interaction. In particular, these trascripts are accepting and the verifier's behavior in them is as in a real interaction. In particular, resorting to the public-coin condition, this means that the verifier's messages in the simulation are (almost) uniformly distributed independently of prior messages.
- In case the input is a NO-instance, the simulator must output either rejecting transcripts or transcripts in which the verifier's behavior is significantly different from the verifier's behavior in a real interaction. In particular, the only way the simulator can produce accepting transcripts is by producing transcripts in which the verifier's messages are not "random enough" (i.e., they depend, in a noticeable way, on previous messages).

Thus assuming, without loss of generality, that the simulator only produces accepting transcripts, we consider two types of distributions. The first type of the distributions is obtained by truncating a random simulator-produced transcript at a random "location" (after some verifier message), whereas the second type is obtained by doing the same while replacing the last verifier message by a random one. Note that both distributions can be implemented by polynomial-size circuits that depend on the input to the proof system being analyzed (and that these two circuits can be constructed in polynomial-time given the said input). The key observation is that if the input is a YES-instance then the two corresponding distributions will be very close, whereas if the input is a NO-instance then there will be a noticeable distance between the two corresponding distributions. Thus, we reduced any problem having a (public-coin) statistical zero-knowledge proof system to $\mathtt{GapSD}^{\mu,\nu}$, where μ is a negligible function and $\nu(n)$ is a noticeable function.[9] The proof is completed by using Theorem 1 (while noting that $\mu(n) < \nu(n)^2 - (1/\mathrm{poly}(n))$).

5.3 Proof sketch for Theorem 3

Theorem 3 was proven by Goldreich and Vadhan [21], by showing that GapCmprEnt is $\mathcal{SZK}$-complete (under Karp-reductions) and invoking Theorem 2 (which shows the same for GapSD). Here we follow a more direct proof, which has appeared in Vadhan's PhD Thesis [22]. The proof consists of two parts: (1) showing that GapSD is Karp-reducible to GapCmprEnt, and (2) showing that GapCmprEnt is Karp-reducible to GapSD.

[9] A function $\mu : \mathsf{N} \to [0,1]$ is called negligible if $\mu(n) < 1/p(n)$ for every positive polynomial p and all sufficiently large n. A function $\nu : \mathsf{N} \to [0,1]$ is called noticeable if $\nu(n) > 1/p(n)$ for some positive polynomial p and all sufficiently large n.

Reducing GapSD to GapCmprEnt: Using Theorem 1, it suffices to reduce $\mathtt{GapSD}^{\epsilon,1-\epsilon}$ to GapCmprEnt, for $\epsilon(n) = 2^{-n}$. Actually, we will reduce $\mathtt{GapSD}^{\epsilon,1-\epsilon}$ to a related problem, denoted GapCmprEnt′, that refers to distinguishing pairs of distributions that have approximately the same entropy from pairs in which the first distribution has (say half a unit of) more entropy.[10] We reduce $\mathtt{GapSD}^{\epsilon,1-\epsilon}$ to GapCmprEnt′ by mapping the circuit pair (C_0, C_1) to (C'_1, C'_2), where $C'_1(r,s,b) \stackrel{\text{def}}{=} (C_s(r), b)$ and $C'_2(r,s,b) \stackrel{\text{def}}{=} (C_s(r), s)$. That is, C_2 outputs a sample of one of the input distributions along with the "selection bit" s used to determine the input distribution being sampled, whereas C_1 outputs such a sample along with an independently distributed random bit (denoted b). Clearly, the entropy of C'_1 is always $v+1$, where $v \stackrel{\text{def}}{=} \frac{\mathrm{H}(C_0)+\mathrm{H}(C_1)}{2}$. Now, if the two input distributions are very far apart then the selection bit s will be determined by the sample and so the entropy of C'_2 will be approximately v, which is significantly smaller than $\mathrm{H}(C'_1)$. On the other hand, if the two input distributions are very close then (even conditioned on the sampled selected) the selection bit s will be almost random and so $\mathrm{H}(C'_2) \approx v+1$, which is approximately the same as $\mathrm{H}(C'_1)$.

A warm-up: reducing GapEnt to GapSD. We first reduce GapEnt to $\mathtt{GapEnt}^{\ell}$, where $\ell(n) = \sqrt{n}$, by using sufficiently many samples (of the input distribution): for example, we may map (C, v) to (C', v'), where $C'(r_1, ..., r_n) = (C(r_1), ..., C(r_n))$ and $v' = n \cdot v$. Next, we assume that the input distribution is "flat", where a distribution is called flat if it is uniform over some set (i.e., if all elements in its support are assigned the same probability mass). We note that by taking sufficiently many samples, we can transform each distribution to one that is "almost flat" (in a sense that is sufficient for the rest of the proof), while maintaining its "relative entropy" (i.e., the average entropy per output bit). Now, suppose that we are given a pair (C, v) such that $C : \{0,1\}^m \to \{0,1\}^n$ is flat and $|\mathrm{H}(C) - v| \geq \sqrt{n}$, and we are interested in the relation between $\mathrm{H}(C)$ and v. Suppose that h is a random hash function[11] mapping m-bit strings to $(m - v - \log_2^2 n)$-bit long string. Now, consider the distributions $(h, C(r), h(r))$ and $(h, C(r), h(r'))$, where $r, r' \in \{0,1\}^m$ and h are uniformly selected. By the property of the hashing function, the third part of the distribution $(h, C(r), h(r'))$ is almost uniform over $\{0,1\}^{m-v-\log_2^2 n}$, even when conditioning on the first parts (specifically on h). On the other hand, the third part of the distribution $(h, C(r), h(r))$ is distributed as $h(r)$ conditioned on $C(r)$ (i.e., $h(r)|C(r)$). We note that $\mathrm{H}(r|C(r)) = m - \mathrm{H}(C)$, and that the distribution $r|C(r)$ is flat. Furthermore, if $\mathrm{H}(C) \leq v$ then $\mathrm{H}(r|C(r)) \geq m - v$ and the distribution $h(r)|C(r)$ is almost uniform over $\{0,1\}^{m-v-\log_2^2 n}$, whereas if $\mathrm{H}(C) \geq v + 2\log_2^2 n$ then $\mathrm{H}(r|C(r)) \leq m - v - 2\log_2^2 n$ and the distribution $h(r)|C(r)$ is very far from being uniform over $\{0,1\}^{m-v-\log_2^2 n}$. Now, recall that $|\mathrm{H}(C) - v| \geq \sqrt{n}$, and observe that if $\mathrm{H}(C) < v$ then the distribution $(h, C(r), h(r))$ is almost identical to the distribution $(h, C(r), h(r'))$, whereas if $\mathrm{H}(C) > v$ then $(h, C(r), h(r))$ is very far from $(h, C(r), h(r'))$. Thus, we have reduced GapEnt to GapSD.

Reducing GapCmprEnt to GapSD: As in the warm-up, we first reduce GapCmprEnt to $\mathtt{GapCmprEnt}^{\ell}$, where $\ell(n) = \sqrt{n}$, such that each of the two distributions is almost flat. Suppose that we are given a pair of circuits (C_1, C_2) such that both are (almost) flat and $|\mathrm{H}(C_1) - \mathrm{H}(C_2)| \geq \sqrt{n}$, and we are interested in the question of which circuit (or distribution represented by it) has higher entropy. Further suppose that $C_1, C_2 : \{0,1\}^m \to \{0,1\}^n$. Suppose that h is a random hash function mapping $(n+m)$-bit strings to $(m - \log_2^2 n)$-bit long string. Now, consider the distributions $(h, C_1(r_1), h(C_2(r_2), r_1))$

[10] Indeed, the reduction from GapCmprEnt′ to GapCmprEnt is easy: we just increase the gap in entropy (by repeated sampling), and move the gap location (by augmenting the second distribution with a few random bits).

[11] Formally speaking, we mean a uniformly selected function in a collection of universal2 hashing functions [8]. For example, we may select h uniformly among all affine mappings of $GF(2^m)$ to $GF(2^k)$, for $k = m - v - \log_2^2 n$.

and $(h, C_1(r_1), h(0^n, r_2))$, where $r_1, r_2 \in \{0,1\}^m$ and h are uniformly selected. By the property of the hashing function, the third part of the distribution $(h, C_1(r_1), h(0^n, r_2))$ is almost uniform over $\{0,1\}^{m-\log_2^2 n}$, even when conditioning on the first parts. On the other hand, the third part of the distribution $(h, C_1(r_1), h(C_2(r_2), r_1))$ is distributed as $h(C_2(r_2), r_1)|C_1(r_1))$. We note that $u \stackrel{\text{def}}{=} \mathrm{H}(C_2(r_2), r_1|C_1(r_1)) = \mathrm{H}(C_2) + (m - \mathrm{H}(C_1))$, and that the distribution $(C_2(r_2), r_1)|C_1(r_1)$ is flat. Furthermore, if $u \geq m$ then the distribution $h(C_2(r_2), r_1)|C_1(r_1))$ is almost uniform over $\{0,1\}^{m-\log_2^2 n}$, whereas if $u \leq m - 2\log_2^2 n$ then the distribution $h(C_2(r_2), r_1)|C_1(r_1))$ is very far from being uniform over $\{0,1\}^{m-\log_2^2 n}$. Now, recall that $|\mathrm{H}(C_1) - \mathrm{H}(C_2)| \geq \sqrt{n}$, and observe that if $\mathrm{H}(C_2) > \mathrm{H}(C_1)$ then $u = m + (\mathrm{H}(C_2) - \mathrm{H}(C_1)) > m$, whereas if $\mathrm{H}(C_2) < \mathrm{H}(C_1)$ then $u \leq m - \sqrt{m}$. We conclose that in the first case the distribution $(h, C_1(r_1), h(C_2(r_2), r_1))$ is almost identical to the distribution $(h, C_1(r_1), h(0^n, r_2))$, whereas in the second case $(h, C_1(r_1), h(C_2(r_2), r_1))$ is very far from $(h, C_1(r_1), h(0^n, r_2))$. Thus, we have reduced `GapCmprEnt` to `GapSD`.

6 Conclusions

In Section 4 we considered the complexity of approximating the entropy of a distribution when given the full description of a sampling device for the distribution. In contrast, the results of Section 3 can be viewed as referring to the case that we are only given "black-box" access to such a sampling device. Thus, the results surveys in these sections represent a potential gap between black-box and "non-black-box" access to sampling devices. This gap may become a real separation if $\mathcal{SZK}$ is contained in sub-exponential time (i.e., $\mathcal{SZK} \subseteq \mathrm{Dtime}(f)$ for some $f(n) = 2^{o(n)}$). On the other hand, the hypothetical existence of "sampling obfuscators" (see [3, Def. 6.2]), which means that non-black-box access to sampling devices does not actually help, implies that $\mathcal{SZK} \neq \mathcal{BPP}$ (see [3, Prop. 6.4]).

We comment that the general study of the relation between black-box and non-black-box algorithms has received considerable attention lately. The interested reader is referred to Barak's PhD Thesis [2].

References

[1] W. Aiello and J. Håstad. Perfect Zero-Knowledge Languages can be Recognized in Two Rounds. In *28th IEEE Symposium on Foundations of Computer Science*, pages 439–448, 1987.

[2] B. Barak. Non-Black-Box Techniques in Cryptography. Ph.D. Thesis, Weizmann Institute of Science, Jan. 2004.

[3] B. Barak, O. Goldreich, R. Impagliazzo, S. Rudich, A. Sahai, S. Vadhan, and K. Yang. On the (im)possibility of software obfuscation. In *Crypto'01*, Springer-Verlag Lecture Notes in Computer Science (Vol. 2139), pages 1–18.

[4] T. Batu, S. Dasgupta, R. Kumar and R. Rubinfeld. The Complexity of Approximating the Entropy. In *34th ACM Symposium on the Theory of Computing*, 2002.

[5] T. Batu, E. Fischer, L. Fortnow, R. Kumar, R. Rubinfeld and P. White. Testing random variables for independence and identity. In *42nd IEEE Symposium on Foundations of Computer Science*, 2001.

[6] T. Batu, L. Fortnow, R. Rubinfeld, W.D. Smith and P. White. Testing that distributions are close. In *41st IEEE Symposium on Foundations of Computer Science*, pages 259–269, 2000.

[7] M. Bellare, O. Goldreich and M. Sudan. Free Bits, PCPs and Non-Approximability – Towards Tight Results. *SIAM Journal on Computing*, Vol. 27, No. 3, pages 804–915, 1998.

[8] L. Carter and M. Wegman. Universal Hash Functions. *Journal of Computer and System Science*, Vol. 18, 1979, pages 143–154.

[9] F. Ergun, S. Kannan, S.R. Kumar, R. Rubinfeld, and M. Viswanathan. Spot-checkers. *Journal of Computer and System Science*, Vol. 60 (3), pages 717–751, 2000.

[10] S. Even, A.L. Selman, and Y. Yacobi. The Complexity of Promise Problems with Applications to Public-Key Cryptography. *Inform. and Control*, Vol. 61, pages 159–173, 1984.

[11] L. Fortnow, The Complexity of Perfect Zero-Knowledge. In *19th ACM Symposium on the Theory of Computing*, pages 204–209, 1987.

[12] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, pages 653–750, July 1998.

[13] O. Goldreich and E. Kushilevitz. A Perfect Zero-Knowledge Proof for a Decision Problem Equivalent to Discrete Logarithm. *Journal of Cryptology*, Vol. 6 (2), pages 97–116, 1993.

[14] O. Goldreich, S. Micali and A. Wigderson. Proofs that Yield Nothing but their Validity or All Languages in NP Have Zero-Knowledge Proof Systems. *Journal of the ACM*, Vol. 38, No. 1, pages 691–729, 1991. Preliminary version in *27th IEEE Symposium on Foundations of Computer Science*, 1986.

[15] O. Goldreich, A. Sahai, and S. Vadhan. Honest-Verifier Statistical Zero-Knowledge equals general Statistical Zero-Knowledge. In *30th ACM Symposium on the Theory of Computing*, pages 399–408, 1998.

[16] O. Goldreich, A. Sahai, and S. Vadhan. Can Statistical Zero-Knowledge be Made Non-Interactive? or On the Relationship of SZK and NISZK. In *Proceedings of Crypto99*, Springer Lecture Notes in Computer Science (Vol. 1666), pages 467–484.

[17] O. Goldreich and S. Vadhan. Comparing Entropies in Statistical Zero-Knowledge with Applications to the Structure of SZK. In *14th IEEE Conference on Computational Complexity*, pages 54–73, 1999.

[18] S. Goldwasser, S. Micali and C. Rackoff. The Knowledge Complexity of Interactive Proof Systems. *SIAM Journal on Computing*, Vol. 18, pages 186–208, 1989. Preliminary version in *17th ACM Symposium on the Theory of Computing*, 1985. Earlier versions date to 1982.

[19] T. Okamoto. On relationships between statistical zero-knowledge proofs. In *28th ACM Symposium on the Theory of Computing*, pages 649–658, 1996.

[20] R. Rubinfeld and M. Sudan. Robust Characterizations of Polynomials with Applications to Program Checking. *SIAM Journal on Computing*, Vol. 25, No. 2, pages 252–271, 1996. Preliminary version in *3rd SODA*, 1992.

[21] A. Sahai and S. Vadhan. A Complete Promise Problem for Statistical Zero-Knowledge. In *38th IEEE Symposium on Foundations of Computer Science*, pages 448–457, 1997.

[22] S. Vadhan. A Study of Statistical Zero-Knowledge Proofs. PhD Thesis, Department of Mathematics, MIT, 1999.

[23] G. Valiant and P. Valiant. A CLT and tight lower bounds for estimating entropy. ECCC, TR10-179, 2010.

[24] G. Valiant and P. Valiant. Estimating the unseen: A sublinear-sample canonical estimator of distributions, ECCC, TR10-180, 2010.

[25] P. Valiant. Testing symmetric properties of distributions. ECCC, TR07-135, 2007.

[26] A. D. Wyner. The wire-tap channel. *Bell System Technical Journal*, Vol. 54 (No. 8), pages 1355–1387, Oct. 1975.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de