Optimal Deterministic Extractors for Generalized Santha-Vazirani Sources

Salman Beigi
Institute for Research in Fundamental Sciences, Tehran, Iran

Andrej Bogdanov
Chinese University of Hong Kong

Omid Etesami
Institute for Research in Fundamental Sciences, Tehran, Iran

Siyao Guo
Northeastern University, Boston, USA

Abstract

Let $\mathcal{F}$ be a finite alphabet and $\mathcal{D}$ be a finite set of distributions over $\mathcal{F}$. A Generalized Santha-Vazirani (GSV) source of type $(\mathcal{F}, \mathcal{D})$, introduced by Beigi, Etesami and Gohari (ICALP 2015, SICOMP 2017), is a random sequence $(F_1, \ldots, F_n)$ in $\mathcal{F}^n$, where F_i is a sample from some distribution $d \in \mathcal{D}$ whose choice may depend on $F_1, \ldots, F_{i-1}$.

We show that all GSV source types $(\mathcal{F}, \mathcal{D})$ fall into one of three categories: (1) non-extractable; (2) extractable with error $n^{-\Theta(1)}$; (3) extractable with error $2^{-\Omega(n)}$.

We provide essentially randomness-optimal extraction algorithms for extractable sources. Our algorithm for category (2) sources extracts one bit with error ε from $n = \mathrm{poly}(1/\varepsilon)$ samples in time linear in n. Our algorithm for category (3) sources extracts m bits with error ε from $n = O(m + \log 1/\varepsilon)$ samples in time $\min\{O(m2^m \cdot n), n^{O(|\mathcal{F}|)}\}$.

We also give algorithms for classifying a GSV source type $(\mathcal{F}, \mathcal{D})$: Membership in category (1) can be decided in NP, while membership in category (3) is polynomial-time decidable.

2012 ACM Subject Classification Theory of computation → Expander graphs and randomness extractors, Mathematics of computing → Probability and statistics, Mathematics of computing → Information theory

Keywords and phrases feasibility of randomness extraction, extractor lower bounds, martingales

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2018.30

Acknowledgements Andrej Bogdanov's work was supported by HK RGC GRF grants CUHK 14208215 and CUHK 14238716. Siyao Guo's work is supported by NSF grants CNS1314722 and CNS-1413964. Part of this work was done while Omid Etesami and Siyao Guo were visiting the Chinese University of Hong Kong, and while Andrej Bogdanov and Siyao Guo were visiting the Simons Institute for the Theory of Computing at UC Berkeley. We would like to thank the reviewers for constructive comments.

1 Introduction

A randomness extractor is an algorithm that converts a weak source of randomness into almost uniform independent random bits. One of the first classes of distributions that were considered in the context of randomness extraction are Santha-Vazirani (SV) sources [16], also called unpredictable-bit sources. An SV source is a sequence of random bits such that every bit in the sequence has entropy bounded away from zero, even when conditioned on

any possible sequence of previous bits. As already pointed out in [16], deterministic (seedless) extraction of even a single almost unbiased bit from SV sources is impossible, although these sources have entropy that grows linearly with their length.[1]

The motivating application for randomness extraction is the simulation of randomized algorithms when only a weak random source is available. For stand-alone algorithms seeded extractors are sufficient to accomplish this simulation. The simulation runs the extractor over all possible seeds and observes the resulting outcomes of the algorithm. In distributed and cryptographic settings, however, multiple executions may be infeasible. In contrast, deterministic extractors of sufficiently high quality are adequate for all these applications.

Even for stand-alone algorithms, simulation using a seeded extractor incurs costs in running time and statistical error. A seed of length ℓ entails 2^ℓ executions of the algorithm and can at best guarantee a statistical error of $2^{-\ell/2}$ [15]. In cryptographic settings, errors are typically inverse-polynomial in the complexity of the adversary, so a simulation of this type may be infeasible. Dodis et al. [9] prove that many cryptographic tasks including encryption, zero-knowledge, secret-sharing, and two-party computation are impossible from arbitrary high-entropy sources including SV sources.

1.1 Generalized Santha-Vazirani sources

In this work we consider deterministic extraction for a natural generalization of Santha-Vazirani sources which was introduced by Beigi, Etesami, and Gohari [2, 3]. A *generalized Santha-Vazirani (GSV) source* is specified by a pair $(\mathcal{F}, \mathcal{D})$, where $\mathcal{F}$ is a finite set of *faces* and $\mathcal{D}$ is a finite set of *dice*, each of which is a probability distribution on $\mathcal{F}$. (We will assume that each face is assigned positive probability by at least one die.) A distribution $(F_1, \ldots, F_n)$, where the F_is are $\mathcal{F}$-valued correlated random variables, is admissible by the source if it is generated by the following type of *strategy*: For each $1 \leq i \leq n$, a die $d \in \mathcal{D}$ is chosen as a function of $F_1, \ldots, F_{i-1}$ and F_i is sampled according to the distribution d. The adversary's choice of the die d may be probabilistic as well (without changing the extractability from the source family).

The case $|\mathcal{F}| = 2$ recovers the definition of SV sources: The dice are two-sided coins, one biased towards heads and the other one towards tails. In this special case the condition $|\mathcal{D}| = 2$ can be imposed without loss of generality by convexity.

▶ **Definition 1.** We call a GSV source $(\mathcal{F}, \mathcal{D})$ *extractable* with error ε from n samples if there exists a function $\mathrm{Ext}\colon \mathcal{F}^n \to \{-1, 1\}$ such that for every distribution $(F_1, \ldots, F_n)$ in the source, $|\mathbb{E}[\mathrm{Ext}(F_1, \ldots, F_n)]| \leq \varepsilon$. We call a source *extractable* if for every error $\varepsilon > 0$ there exists a sample size n for which the source is extractable with these parameters.

The work [3] showed that randomness extraction from a GSV source is possible assuming the following condition:

▶ **Definition 2.** A GSV source $(\mathcal{F}, \mathcal{D})$ satisfies the *Nonzero Kernel Positive Variance (*NK^+*)* condition if there exists a function $\psi\colon \mathcal{F} \to [-1, 1]$ such that $\mathbb{E}_d[\psi(F)] = 0$ and $\mathrm{Var}_d[\psi(F)] > 0$ for every die $d \in \mathcal{D}$.

Here, $\mathbb{E}_d$ and Var_d denote expectation and variance with respect to the distribution of die d. On the other hand, they showed that extractability from such sources necessitates the following *Nonzero Kernel* (NK) condition:

There exists a nonzero $\psi\colon \mathcal{F} \to [-1, 1]$ such that $\mathbb{E}_d[\psi(F)] = 0$ for every die $d \in \mathcal{D}$.

[1] With respect to seeded extraction, a constant seed length is sufficient for all SV sources [17].

In particular, when all faces of all dice have positive probability (an assumption called "nondegeneracy" in [3]), the (NK^+) and (NK) conditions coincide, providing a characterization of extractability for this class of sources. Their extractor requires $\Theta(1/\varepsilon^3)$ samples to achieve error ε.

There are, however, simple examples of GSV sources ((E1) and (E2) below) that satisfy (NK) but not (NK^+). The work [3] does not address the extractability of such sources.

The existence of extractors for GSV sources does not appear to easily follow from counting arguments, as is the case of other types of sources for which extraction is known to be possible in principle and the focus is on efficient constructions, such as affine sources [6, 12], polynomial sources [11, 10] and independent blocks [5, 7].

1.2 Our Contributions

Our first contribution is a complete characterization of extractability from GSV sources. To motivate our result, we first observe that the (NK) condition is, in general, insufficient for extractability. Consider, for instance the two-diced, three-faced GSV source described by the distributions (i.e., probability mass functions) $d_1 = (0, 0, 1)$ and $d_2 = (\frac{1}{2}, \frac{1}{2}, 0)$. This source satisfies (NK) with the witness $\psi = (-1, 1, 0)$, but is clearly not extractable as the distribution in which d_1 is repeatedly tossed contains no entropy.

The following GSV source is a slightly more interesting example:

$$d_1 = (\tfrac{1}{2}, \tfrac{1}{2}, 0, 0) \quad d_2 = (0, 0, \tfrac{1}{3}, \tfrac{2}{3}) \quad d_3 = (0, 0, \tfrac{2}{3}, \tfrac{1}{3}). \tag{E1}$$

This source also satisfies the (NK) condition (with $\psi = (-1, 1, 0, 0)$). However, it is not extractable because it contains a "hidden" SV source (over two faces): If die d_1 is discarded and the first two faces are removed, dice d_2 and d_3 now fail the (NK) condition.

These two examples suggest the following method for coming up with non-extractable GSV sources: Start with any source that fails (NK), extend the dice with more faces of zero probability, and add any number of dice that assign positive probability to the new faces. To describe such sources, we introduce the following natural strengthening of (NK):

▶ **Definition 3.** A GSV source $(\mathcal{F}, \mathcal{D})$ satisfies the *Hereditary Nonzero Kernel (HNK)* condition if for every subset $\mathcal{D}' \subseteq \mathcal{D}$ there exists a nonzero function $\psi : \mathcal{F}' \to [-1, 1]$ such that $\mathbb{E}_d[\psi(F)] = 0$ for all $d \in \mathcal{D}'$, where $\mathcal{F}' = \mathcal{F}'(\mathcal{D}')$ is the set of faces to which at least one die in $\mathcal{D}'$ assigns nonzero probability.

Clearly (HNK) is a necessary condition for extractability, because if $(\mathcal{F}, \mathcal{D})$ fails (HNK) then $(\mathcal{F}', \mathcal{D}')$ fails (NK). Our first theorem shows that (HNK) is also sufficient. Moreover, it gives a universal upper bound on the number of samples:

▶ **Theorem 4.** *The following conditions are equivalent for a GSV source* $(\mathcal{F}, \mathcal{D})$*:*

1. $(\mathcal{F}, \mathcal{D})$ *satisfies HNK.*
2. $(\mathcal{F}, \mathcal{D})$ *is extractable.*
3. *For every* ε*,* $(\mathcal{F}, \mathcal{D})$ *is extractable with error* ε *from* $n = \mathrm{poly}(1/\varepsilon)$ *samples in time linear in* n*.*

In the course of proving Theorem 4 we introduce the analytic *Mean Variance Ratio (MVR)* condition that turns out to be equivalent to HNK (Proposition 12). We show that a quantitative variant of the MVR condition determines the best-possible quality of extraction, up to a quadratic gap, even for GSV sources that are not extractable to within arbitrary small error (Propositions 7 and 10).

The work [3] proved statement 3 of Theorem 4 for the subclass of NK^+ sources. Theorem 5 shows that NK^+ are in fact extractable from a logarithmic number of samples, and they are the only sources for which this degree of efficiency is possible.

▶ **Theorem 5.** *The following conditions are equivalent for a GSV source $(\mathcal{F}, \mathcal{D})$:*

1. *$(\mathcal{F}, \mathcal{D})$ satisfies NK^+.*
2. *For every ε, $(\mathcal{F}, \mathcal{D})$ is extractable with error ε from $o(1/\varepsilon^2)$ samples.*
3. *For every ε and m, $(\mathcal{F}, \mathcal{D})$ is extractable with error[2] ε and output length m from $n = O(\log(1/\varepsilon) + m)$ samples in time $\min\{O(m2^m \cdot n), n^{O(|\mathcal{F}|)}\}$.*

The sample complexity of the extractor in part 3 of Theorem 5 is optimal up to the leading constant: $\Omega(m)$ samples are necessary by entropy considerations, and $\Omega(1/\varepsilon)$ samples are necessary for non-trivial sources[3] by granularity considerations.

Condition NK^+ is strictly stronger than condition HNK. For example, the source

$$d_1 = (\tfrac{1}{2}, \tfrac{1}{2}, 0, 0), \quad d_2 = (\tfrac{1}{4}, \tfrac{1}{12}, \tfrac{1}{3}, \tfrac{1}{3}), \quad d_3 = (\tfrac{1}{12}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{1}{3}). \tag{E2}$$

satisfies HNK but not NK^+.

Taken together, Theorems 4 and 5 completely classify non-trivial GSV sources into three categories: (1) non-extractable, (2) extractable with error $n^{-\Theta(1)}$, and (3) extractable with error $2^{-\Omega(n)}$, where n is the number of samples. This rules out the existence of GSV sources of other error rates like $1/\log n$ or $2^{-\sqrt{n}}$.

Moreover, sources can be classified algorithmically: Condition HNK can be decided by a coNP algorithm, while NK^+ is polynomial-time decidable (see Proposition 18).

Figure 1 indicates the relations between the different conditions for extractability of GSV sources uncovered in this work. The left and right columns describe equivalent formulations of randomness-efficient and general extractability, respectively. The middle column contains the different types of GSV sources in increasing generality from top to bottom. The HNK condition is shown equivalent to all formulations of general extractability, while the NK^+ condition [3] is shown equivalent to randomness-efficient extractability.

1.3 Proof Techniques

Feasibility of extraction. The extractor of [3] outputs the sign of $Z_T = \psi(F_1) + \cdots + \psi(F_T)$ at the earliest time T when $|Z_T|$ exceeds some pre-specified threshold M. Here, ψ is the witness for condition (NK^+), which ensures that $\mathbb{E}[\psi(F)]$ is always zero and $\mathrm{Var}[\psi(F)]$ is always positive. Therefore (Z_t) is a martingale with growing variance, and the analysis of [3] shows that the process terminates by time $n = O(1/\varepsilon^3)$ except with probability $\varepsilon/2$ when M is chosen as $\Theta(1/\varepsilon)$. Moreover, Z_T must take value in the range $(-(M+1), -M] \cup [M, M+1)$, so by the optional stopping time theorem, the bias of Z_T is $\varepsilon/2$ when $M = \Theta(1/\varepsilon)$.

In case only the weaker (HNK) condition holds, $\mathrm{Var}[\psi(F)]$ could be zero for some dice and the value of Z_t may remain constant throughout the process. On the other hand, (HNK) provides not one but many witnesses ψ, one for every subset of the dice. Proposition 12 shows how all these witnesses can be combined into a single $\phi\colon \mathcal{F} \to [-1, 1]$ that has positive variance with respect to all the dice, but may have nonzero expectation. By a careful

[2] The error of an extractor that outputs multiple bits is the statistical (total variation) distance between its output distribution and the uniform distribution.

[3] The exception consists of GSV sources that admit an event of probability exactly half after throwing only one die, for which errorless extraction is possible.

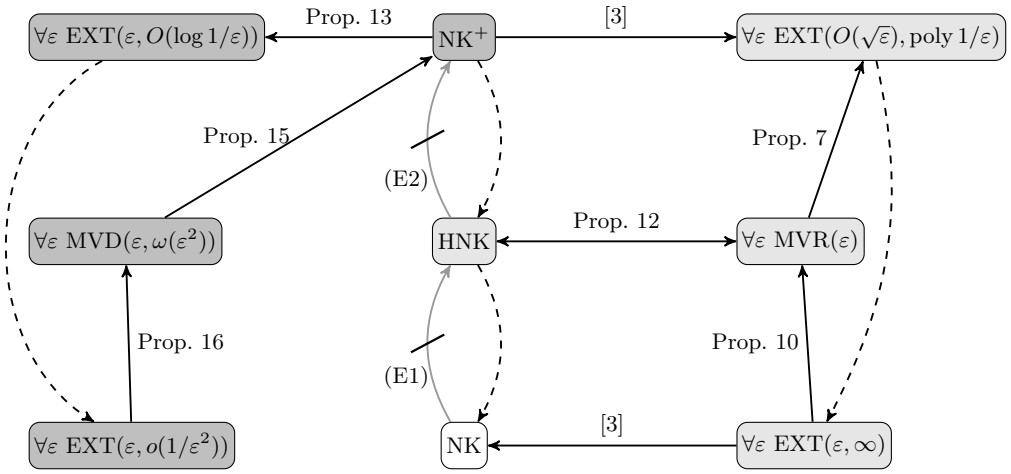

Figure 1 A map of our results. Straight arrows are implications (the dashed ones are immediate) and lighter struck-out arrows are separations given by Examples (E1) and (E2). EXT(ε, n) postulates extractability with error ε from n samples. Lightly and darkly shaded boxes represent equivalent conditions for extractability and randomness-efficient extractability, respectively. Definitions 2, 3, 6, and 14 specify the NK$^+$, HNK, MVR, and MVD conditions, respectively.

implementation of this strategy, it is ensured that the ratio $|\mathbb{E}_d[\phi(F)]|/\mathrm{Var}_d[\phi(F)]$ can be made smaller than any pre-specified $\varepsilon > 0$. This is our Mean Variance Ratio (MVR) condition. Moreover, $\mathrm{Var}_d[\phi(F)]$ can be lower bounded by ε^C for some constant C that depends only on the GSV source.

To prove Theorem 4 we apply the extractor of [3] to the function ϕ. As ϕ may be biased with respect to some dice, (Z_t) may no longer be a martingale, rendering the optional stopping time theorem inapplicable. In Proposition 7 we demonstrate that the conclusion of the [3] analysis still applies in our context. Intuitively, the (MVR) condition should imply that the variance of Z_t grows, and does so at a faster rate than the magnitude of its expectation. Therefore the stopping time should still be finite, and the component of extraction error incurred by $|\mathbb{E}[Z_T]|$ should be small. Owing to dependencies between the various steps, a rigorous implementation of these ideas requires substantial care.

Quality and quantity of extracted bits. For GSV sources that satisfy (NK$^+$) the extractor of [3] inherently requires $\Omega(1/\varepsilon)$ samples: On the one hand, to ensure termination with high probability the boundary threshold M can be at most n, but on the other hand Z_T may fall anywhere in the range $(-(M+1), -M] \cup [M, M+1)$, thereby incurring an error of $\varepsilon = \Omega(1/M)$.[4] To improve the sample complexity, our bit extractor in Theorem 5 applies the update rule

$$Z_{t+1} = Z_t + \frac{\psi(F_t)}{2} \cdot (1 - |Z_t|)$$

and outputs the sign of Z_n for $n = O(\log 1/\varepsilon)$. Under (NK$^+$) the sequence (Z_t) is still a martingale, but now the range of Z_t is restricted to the open interval $(-1, 1)$. On average, the deviation of the step size $Z_{t+1} - Z_t$ conditioned on Z_t is smaller the closer Z_t is to one

[4] A tempting alternative is to simply output the sign of Z_n after looking at some predetermined number of samples. However, this "extractor" incurs error $\Omega(1)$ for any non-trivial GSV source.

of the boundary points $\{-1, 1\}$. We show that the logarithm of $1/(1 - |Z_t|)$ grows by a constant on average in every step and apply Azuma's inequality to conclude that Z_n is within $2^{-\Omega(n)}$ of 1 or -1 with probability $1 - 2^{-\Omega(n)}$. This ensures the bias of the output is inverse exponential in the number of samples.

We give some informal intuition about the fast convergence of the martingale. Since its range is bounded to $[-1, 1]$ and it is "unstable" at any point far from the boundaries, by Doob's theorem it must converge to either -1 or 1. The extractor's decision about its output should not be influenced heavily by the last few symbols of the source. Otherwise, the extractor can be biased by an adversarial choice of distribution. (In particular, a random function is typically a poor extractor since the last bits have very high influence.) A "sufficient statistic" for the extractor's future strategy at any point is the bias of its output given the symbols read so far. This probability is not a well-defined quantity since the source is selected adversarially. The (NK$^+$) condition enables a rigorous (and optimal) implementation of this strategy.

To extract multiple bits, the state $\mathbf{Z}_t$ of the above process is extended to encode a probability distribution over $\{0, 1\}^m$. Initially $\mathbf{Z}_0$ is the uniform distribution. The distance measure $1 - |Z_t|$ is replaced by a carefully chosen quantity $\mathbf{D}_t \in \mathbb{R}^{2^m}$ which ensures that $\mathbf{Z}_t$ is a probability distribution that rapidly concentrates on a single entry in $\{0, 1\}^m$, which is the output of the extractor. Since $(\mathbf{Z}_t)$ is a multi-dimensional martingale, the output must be statistically close to uniform. The straightforward method of keeping track of the corresponding 2^m martingales requires exponential space, but we show a way to maintain a succinct representation of them.

Lower bounds. Beigi, Etesami, and Gohari [3] proved that if a source fails the (NK) condition, then it is not extractable. There are two proofs of this fact in [3], an analytic one (in Section 2.2) and a combinatorial one (in Appendix B). Here we refine the ideas of the analytic proof. In Proposition 10 we prove a quantitatively precise refinement of this statement: As we shall show, the (NK) condition fails if for all ψ there exists a die d for which $|\mathbb{E}_d[\psi(F)]|/\mathrm{Var}_d[\psi(F)] = \Omega(1)$. Now if $|\mathbb{E}_d[\psi(F)]|/\mathrm{Var}_d[\psi(F)] \geq \varepsilon$, then the extraction error must be at least $\Omega(\varepsilon)$. We conclude that extractability implies the (MVR) condition, which together with a compactness argument (see Proposition 12) gives (HNK), proving the "only if" direction of Theorem 4.

We prove the "only if" direction of Theorem 5 in Section B. We introduce the mean-variance divergence (MVD) condition, which postulates that $|\mathbb{E}_d[\psi(F)]| < \varepsilon(\mathrm{Var}_d[\psi(F)] - \delta)$ for all dice. In Proposition 16 we show that if MVD fails then extraction with error ε requires $\Omega(1/\delta)$ samples. In Proposition 15 we use linear-algebraic duality to show that if (NK$^+$) fails then so does (MVD) with $\delta = O(\varepsilon^2)$, thereby completing the proof of Theorem 5.

1.4 Other related work

The question of extractability from GSV sources has found applications in tampering attacks in cryptography [14, 13] and in publicly verifiable randomness via cryptocurrencies [4].

The GSV sources also capture the block sources of [8]. The latter (ℓ, b)-sources correspond to the special case of $\mathcal{F} = \{0, 1\}^\ell$ and $\mathcal{D}$ containing all flat distributions over $\mathcal{F}$ having min-entropy at least b. Thus our theorems refine the impossibility of deterministic extraction from block sources.

2 A characterization of extractable GSV sources

In this Section we prove Theorem 4. The following analytic condition plays a central role in the proof:

▶ **Definition 6.** A GSV source $(\mathcal{F}, \mathcal{D})$ satisfies the *Mean-Variance Ratio condition* with parameter $\varepsilon > 0$ (MVR(ε)) if there exists a function $\psi : \mathcal{F} \to [-1, 1]$ such that for every die $d \in \mathcal{D}$ of a GSV source $(\mathcal{F}, \mathcal{D})$,

$$\left|\mathbb{E}_d[\psi(F)]\right| < \epsilon \mathrm{Var}_d[\psi(F)]. \tag{MVR}$$

Proposition 7 in Section 2.1 shows that if a GSV source satisfies MVR(ε) then it is extractable with error $O(\sqrt{\varepsilon})$ from poly$(1/\varepsilon)$ samples. On the other hand, Proposition 10 in Section 2.2 shows that any GSV source that is extractable with error less than $\varepsilon/10$ (from any number of samples) satisfies MVR(ε). Thus the smallest ε for which MVR(ε) holds measures the best-possible quality of extraction of a GSV source to within a square.

In the case when MVR(ε) holds for all $\varepsilon > 0$, the source is extractable. Proposition 12 shows that if this is the case then HNK must hold. HNK, in turn, implies a slightly stronger form of "MVR(ε) for all ε". Together with Proposition 7 this establishes the extractability of HNK sources from ε^{-C} samples, where C is a constant that depends only on the source.

2.1 Feasibility of extraction

▶ **Proposition 7.** *If GSV source $(\mathcal{F}, \mathcal{D})$ satisfies* MVR(ε)*, then it is extractable from n samples with error at most $3\sqrt{\varepsilon} + 4/\varepsilon v n + O(\varepsilon)$, where v is the minimum of* $\mathrm{Var}_d[\psi(F)]$ *over all $d \in \mathcal{D}$.*

The extractor outputs the sign of $\psi(F_1) + \cdots + \psi(F_T)$ if $T \le n$, where F_i is the i-th output of the GSV source sequence, and T is the first time when the magnitude of this expression exceeds the value $M = 1/\sqrt{\varepsilon}$. The output can be arbitrary if $T > n$.

The sequence $X_t = \psi(F_t)$ is a special case of an (ε, v)-*approximate martingale*, namely a sequence (X_t), $|X_t| \le 1$ such that

$$|\mathbb{E}[X_t \mid X_{<t}]| < \varepsilon \cdot \mathrm{Var}[X_t \mid X_{<t}] \qquad \text{and} \qquad \mathrm{Var}[X_t \mid X_{<t}] \ge v$$

for all t and $X_{<t} = (X_1, \ldots, X_{t-1})$. The key result is the following lemma which bounds the deviation time of an approximate martingale and its bias at the deviation time.

▶ **Lemma 8.** *Let $Z_t = X_1 + \cdots + X_t$ and T be the first time when $|Z_T| \ge M$ for an (ε, v)-approximate martingale (X_t). Then $|\mathbb{E}[Z_T]| \le \varepsilon(M+1)^2$ and $\Pr[T > t] \le 2(M+1)^2/vt$, assuming $\varepsilon \le 1/8(M+1)$.*

Proof of Lemma 8. We modify the sequence so that $X_t = 0$ for all $t > T$. The variables of interest T and Z_T do not change, while the approximate martingale assumptions imply

$$|\mathbb{E}[X_t \mid X_{<t}]| \le \varepsilon \cdot \mathrm{Var}[X_t \mid X_{<t}] \qquad \text{and} \qquad \mathrm{Var}[X_t \mid X_{<t}, T > t] \ge v.$$

The advantage of this modification is that $|Z_t|$ is now upper bounded by $M+1$ for all t. The upper bounds on $\Pr[T > t]$ and $|\mathbb{E}[Z_T]|$ will both follow from this variance lower bound:

▶ **Claim 9.** *For all t,* $\mathrm{Var}[Z_t] \ge \frac{1}{2}\sum_{i=1}^{t} \mathbb{E}[\mathrm{Var}[X_i | X_{<i}]]$.

Proof. By the law of total variance, for every t,

$$\mathrm{Var}[Z_t] = \mathrm{Var}\big[\mathbb{E}[Z_t|X_{<t}]\big] + \mathbb{E}\big[\mathrm{Var}[Z_t|X_{<t}]\big].$$

Furthermore,

$$\begin{aligned}\mathrm{Var}\big[\mathbb{E}[Z_t|X_{<t}]\big] &= \mathrm{Var}\big[Z_{t-1} + \mathbb{E}[X_t|X_{<t}]\big] \\ &= \mathrm{Var}[Z_{t-1}] + \mathrm{Var}\big[\mathbb{E}[X_t|X_{<t}]\big] + 2\mathrm{Cov}\big(Z_{t-1}, \mathbb{E}[X_t|X_{<t}]\big).\end{aligned}$$

We can lower bound the covariances by

$$\begin{aligned}\mathrm{Cov}\big(Z_{t-1}, \mathbb{E}[X_t|X_{<t}]\big) &= \mathbb{E}\big[\,(Z_{t-1} - \mathbb{E}[Z_{t-1}]) \cdot \mathbb{E}[X_i|X_{<t}]\,\big] \\ &\geq -\mathbb{E}\big[\,|Z_{t-1} - \mathbb{E}[Z_{t-1}]| \cdot |\mathbb{E}[X_t|X_{<t}]|\,\big] \\ &\geq -\mathbb{E}\big[\,(2M+2) \cdot \epsilon \mathrm{Var}[X_t|X_{<t}]\,\big] \\ &\geq -\frac{1}{4}\mathbb{E}\big[\mathrm{Var}[X_t|X_{<t}]\big],\end{aligned}$$

where the penultimate inequality follows from the boundedness of Z_{t-1} and the last one follows from the assumption $\varepsilon \leq 1/8(M+1)$.

Combining the above (in)equalities and using the nonnegativity of $\mathrm{Var}[\mathbb{E}[X_t|X_{<t}]]$,

$$\begin{aligned}\mathrm{Var}[Z_t] &\geq \mathbb{E}\big[\mathrm{Var}[Z_t|X_{<t}]\big] + \mathrm{Var}[Z_{t-1}] - \frac{1}{2}\mathbb{E}\big[\mathrm{Var}[X_t|X_{<t}]\big] \\ &= \mathrm{Var}[Z_{t-1}] + \frac{1}{2}\mathbb{E}\big[\mathrm{Var}[X_t|X_{<t}]\big].\end{aligned}$$

The claim now follows by induction on t. ◀

We can now upper bound $|\mathbb{E}[Z_T]|$ by the maximum of $\mathbb{E}[Z_t]$ over all t, which is at most

$$|\mathbb{E}[Z_t]| \leq \sum_{i=1}^{t} \mathbb{E}\big[|\mathbb{E}[X_i|X_{<i}]|\big] \leq \sum_{i=1}^{t} \mathbb{E}\big[\varepsilon \mathrm{Var}[X_i|X_{<i}]\big] \leq 2\varepsilon \mathrm{Var}[Z_t] \leq 2\varepsilon(M+1)^2. \tag{1}$$

The second inequality follows from the approximate martingale assumption. The third one follows from Claim 9. The fourth one follows from the boundedness of Z_t.

It remains to upper bound $\Pr[T > t]$. By Claim 9, the law of conditional expectations, and the approximate martingale assumption,

$$\begin{aligned}\mathrm{Var}[Z_t] &\geq \sum_{i=1}^{t} \mathbb{E}\big[\mathrm{Var}[X_i|X_{<i}]\big] \\ &= \frac{1}{2}\sum_{i=1}^{t} \Pr[T > i] \cdot \mathbb{E}\big[\mathrm{Var}[X_i|X_{<i}] \,\big|\, T > i\big] \\ &\geq \frac{1}{2}\sum_{i=1}^{t} \Pr[T > t] \cdot v \\ &= \frac{tv}{2} \cdot \Pr[T > t].\end{aligned}$$

The desired lower bound follows from the boundedness of Z_t. ◀

Proof of Proposition 7. Let $X_t = \psi(F_t)$, $Z_t = X_1 + \cdots + X_t$ and T be the first time when $|Z_T| \geq M$. Conditioned on $T \leq n$, the output Ext of the extractor is identically distributed to $\mathrm{sign}(Z_T)$. Therefore by the law of conditional expectations,

$$|\mathbb{E}[\mathrm{Ext}] - \mathbb{E}[\mathrm{sign}(Z_T)]| = |\mathbb{E}[Ext \mid T > n] - \mathbb{E}[Z_T \mid T > n]| \cdot \Pr[T > n] \leq 2\Pr[T > n].$$

By the boundedness of the X_ts,

$$|Z_T - M \cdot \mathrm{sign}(Z_T)| \leq 1.$$

By the triangle inequality and Lemma 8,

$$\begin{aligned}|\mathbb{E}[\mathrm{Ext}]| &\leq |\mathbb{E}[\mathrm{sign}(Z_T)]| + 2\Pr[T > n] \\ &\leq \frac{|\mathbb{E}[Z_T]| + 1}{M} + 2\Pr[T > n] \\ &\leq \frac{\varepsilon(M+1)^2}{M} + \frac{1}{M} + \frac{4(M+1)^2}{vn},\end{aligned}$$

assuming $\varepsilon \leq 1/8(M+1)$. When $M = 1/\sqrt{\varepsilon}$ the assumption is satisfied for every $\varepsilon \leq 1/4$ and the desired bound follows. When $\varepsilon > 1/4$ the claimed bias is larger than one and there is nothing to prove. ◀

2.2 Impossibility of extraction

▶ **Proposition 10.** *Let ε be a sufficiently small constant. Assume* $\mathrm{MVR}(\varepsilon)$ *fails for a source $(\mathcal{F}, \mathcal{D})$. Then $(\mathcal{F}, \mathcal{D})$ is not extractable with error better than $\varepsilon/10$ from any number of samples.*

Proof of Proposition 10. Assuming $\mathrm{MVR}(\varepsilon)$ fails we prove the following claim:

▶ **Claim 11.** *For every n, every extractor* $\mathrm{Ext}\colon \mathcal{F}^n \to \{0,1\}$*, and every $0 \leq \alpha \leq 1$, if $\mathbb{E}_{A_-}[\mathrm{Ext}] \geq \alpha$ for every strategy A_-, then there exists a strategy A_+ for which $\mathbb{E}_{A_+}[\mathrm{Ext}] \geq \alpha + (\varepsilon/(1+\varepsilon)) \cdot \alpha(1-\alpha)$.*

To derive the theorem from the claim, assume that $\mathbb{E}[\mathrm{Ext}] \geq \alpha = 1/2 - \varepsilon/10$ with respect to every strategy. By Claim 11 there must then exist a strategy for which

$$\mathbb{E}[\mathrm{Ext}] \geq \frac{1}{2} - \frac{\varepsilon}{10} + \frac{\varepsilon}{1+\varepsilon} \cdot \frac{1 - \varepsilon^2/100}{4}$$

which is at least $1/2 + \varepsilon/10$. ◀

Beigi, Etesami, and Gohari prove Claim 11 under the stronger assumption that NK fails. Their analysis proves the lower bound $\mathbb{E}_{A_+}[\mathrm{Ext}] \geq \alpha + \varepsilon f(\alpha)$ for a more general class of functions f [3, Lemma 10]. The form that we choose is convenient for balancing the simultaneous requirements on mean and variance.

Proof of Claim 11. We prove the claim by induction on n. When $n = 0$ the claim holds by checking the cases $\mathrm{Ext} = 0$ and $\mathrm{Ext} = 1$. We now assume it holds for $n-1$ and prove it for n. Fix A_- to be the strategy that minimizes $\mathbb{E}_{A_-}[\mathrm{Ext}]$, that is at least α. Let d_- be the choice of the first die in this strategy. Then

$$\alpha \leq \mathbb{E}_{d_-}[\alpha_F],$$

where α_f is the conditional expectation of Ext given the first outcome being f, under strategy A_-.

We now describe the strategy A_+. By $\overline{\mathrm{MVR}(\varepsilon)}$ applied to the function $\psi(f) = \alpha_f - \alpha$, there exists a die d_+ such that

$$\mathbb{E}_{d_+}[\alpha_F - \alpha] \geq \varepsilon \mathrm{Var}_{d_+}[\alpha_F]. \tag{2}$$

The adversary A_+ tosses this die first. She then plays the strategy that maximizes $\mathbb{E}_{A_+}[\mathrm{Ext}]$ conditioned on the outcome of the first die. By our inductive assumption, the conditional expectation of Ext when A_+ is played, given the first outcome is f, must be at least $\alpha_f + (\varepsilon/(1+\varepsilon)) \cdot \alpha_f(1-\alpha_f)$ so that

$$\mathbb{E}_{A_+}[\mathrm{Ext}] \geq \mathbb{E}_{d_+}\Big[\alpha_F + \frac{\varepsilon}{1+\varepsilon} \cdot \alpha_F(1-\alpha_F)\Big].$$

We can write

$$\begin{aligned}\mathbb{E}_{d_+}\Big[\alpha_F + \frac{\varepsilon}{1+\varepsilon} \cdot \alpha_F(1-\alpha_F)\Big] - \Big(\alpha + \frac{\varepsilon}{1+\varepsilon} \cdot \alpha(1-\alpha)\Big) \\ = \Big(1 + \frac{\varepsilon}{1+\varepsilon}\Big)\mathbb{E}_{d_+}[\alpha_F - \alpha] - \frac{\varepsilon}{1+\varepsilon}\mathrm{Var}_{d_+}[\alpha_F] - \frac{\varepsilon}{1+\varepsilon}\big(\mathbb{E}_{d_+}[\alpha_F]^2 - \alpha^2\big). \quad (3)\end{aligned}$$

For the last term we have the upper bound

$$\mathbb{E}_{d_+}[\alpha_F]^2 - \alpha^2 = \mathbb{E}_{d_+}[\alpha_F + \alpha] \cdot \mathbb{E}_{d_+}[\alpha_F - \alpha] \leq 2\mathbb{E}_{d_+}[\alpha_F - \alpha],$$

since all the αs are between zero and one, and the second term is non-negative because $\mathbb{E}_{d_+}[\alpha_F] \geq \mathbb{E}_{d_-}[\alpha_F] \geq \alpha$ by the minimality of d_-. We can therefore lower bound the left hand size of (3) by

$$\Big(1 - \frac{\varepsilon}{1+\varepsilon}\Big)\mathbb{E}_{d_+}[\alpha_F - \alpha] - \frac{\varepsilon}{1+\varepsilon}\mathrm{Var}_{d_+}[\alpha_F].$$

By (2) this must be non-negative. It follows that $\mathbb{E}_{A_+}[\mathrm{Ext}]$ is at least $\alpha + (\varepsilon/(1+\varepsilon))\alpha(1-\alpha)$, concluding the inductive step. ◀

2.3 Proof of Theorem 4

▶ **Proposition 12.** *The following conditions are equivalent for a GSV source $(\mathcal{F}, \mathcal{D})$:*

1. *For all $\varepsilon > 0$, $(\mathcal{F}, \mathcal{D})$ satisfies* MVR(ε)*: There exists a $\psi : \mathcal{F} \to [-1, 1]$ such that for all dice d, $|\mathbb{E}_d[\psi(F)]| < \varepsilon \mathrm{Var}_d[\psi(F)]$.*
2. *There exists a constant C such that for sufficiently small $\varepsilon > 0$, there exists a $\psi : \mathcal{F} \to [-1, 1]$ such that for all dice d, $|\mathbb{E}_d[\psi(F)]| < \varepsilon \mathrm{Var}_d[\psi(F)]$ and $\mathrm{Var}_d[\psi(F)] \geq \varepsilon^C$.*
3. *$(\mathcal{F}, \mathcal{D})$ satisfies* HNK*.*

Proof. We will show that 1 implies 3 and 3 implies 2. This will establish equivalence as 2 is a stronger condition than 1.

1 implies 3: Assume that $(\mathcal{F}, \mathcal{D})$ satisfies MVR(ε). This condition is hereditary, namely if it holds for $(\mathcal{F}, \mathcal{D})$ then it holds for all $(\mathcal{F}', \mathcal{D}')$ in the assumption of HNK. So in proving 3, we may and will assume, without loss of generality, that $(\mathcal{F}', \mathcal{D}') = (\mathcal{F}, \mathcal{D})$. We will moreover assume (by scaling and flipping sign if necessary) that ψ attains the value 1.

Now consider an infinite decreasing sequence (ε_k) that converges to zero. By assumption, for every k there exists a ψ_k such that $|\mathbb{E}_d[\psi_k(F)]| < \varepsilon_k \mathrm{Var}_d[\psi_k(F)]$. By the pigeonhole principle there must exist a face f for which the set of indices $K = \{k \colon \psi_k(f) = 1\}$ is infinite. By compactness of $[-1, 1]^{\mathcal{F}}$ there must exist an infinite subset $K' \subseteq K$ for which the subsequence ψ_k over $k \in K'$ converges to a limit ψ. Then ψ is nonzero as $\psi(f)$ must equal one. On the other hand, for every $\varepsilon > 0$ there exists a sufficiently large $k \in K'$ such that for every die d,

$$|\mathbb{E}_d[\psi(F)]| \leq |\mathbb{E}_d[\psi_k]| + \varepsilon \leq \varepsilon \mathrm{Var}_d[\psi_k(F)] + \varepsilon,$$

so $\mathbb{E}_d[\psi(F)]$ must equal zero for every d.

3 implies 2: The proof is by strong induction on the number of dice $|\mathcal{D}|$ with $C = 3 \cdot 2^{|\mathcal{D}|} - 3$. In the base case $|\mathcal{D}| = 1$, all faces must be assigned nonzero probability by the unique die d. Take any witness ψ for HNK. Then $\mathbb{E}_d[\psi(F)] = 0$, but ψ must take nonzero value on at least one of the faces, so $\mathrm{Var}_d[\psi(F)] > 0$. Condition 2 is then satisfied for sufficiently small $\varepsilon > 0$.

For the inductive step, take any ψ that is a witness for HNK with respect to the whole source $(\mathcal{F}, \mathcal{D})$. Let $\mathcal{D}'$ be the subset of dice d such that $\mathrm{Var}_d[\psi(F)] = 0$ and v be the minimum of $\mathrm{Var}_d[\psi(F)]$ over $d \notin \mathcal{D}'$. Then $\mathcal{D}'$ is a proper subset of $\mathcal{D}$ (otherwise, there is a face that is assigned no probability by any die). If $\mathcal{D}'$ is empty, condition 2 follows by the same argument as in the base case. If not, then by the inductive hypothesis we can choose $\psi' \colon \mathcal{F}' \to [-1, 1]$ such that

$$|\mathbb{E}_d[\psi'(F)]| < (v\varepsilon^2/8) \cdot \mathrm{Var}_d[\psi(F)] \quad \text{and} \quad \mathrm{Var}_d[\psi(F)] \geq (v\varepsilon^2/8)^{3 \cdot 2^{|\mathcal{D}'|} - 3}. \tag{4}$$

We will show that the function $\phi = \psi + (v\varepsilon/8) \cdot \psi'$ satisfies the conclusion of condition 2. Here, ψ' is naturally extended as a function on $\mathcal{F}$ by assigning zero on all inputs in $\mathcal{F} \setminus \mathcal{F}'$. The proof is by cases.

If $d \in \mathcal{D}'$, then $\mathbb{E}_d[\phi(F)] = (v\varepsilon/8)\mathbb{E}_d[\psi'(F)]$, while $\mathrm{Var}_d[\phi(F)] = (v\varepsilon/8)^2 \mathrm{Var}_d[\psi'(F)]$. From these two equalities and (4) it follows that $\mathbb{E}_d[\phi(F)] < \varepsilon \mathrm{Var}_d[\phi(F)]$. On the other hand, $\mathrm{Var}_d[\phi(F)] \geq (v\varepsilon/8)^2 \cdot (v\varepsilon^2/8)^{3 \cdot 2^{|\mathcal{D}'|} - 3} \geq \varepsilon^{3 \cdot 2^{|\mathcal{D}|} - 3}$ for sufficiently small ε.

If $d \notin \mathcal{D}'$, then $|\mathbb{E}_d[\phi(F)]| \leq (v\varepsilon/8)|\mathbb{E}_d[\psi'(F)]| \leq v\varepsilon/8$, while

$$\begin{aligned}
\mathrm{Var}_d[\psi(F)] &\geq \mathrm{Var}_d[\psi'(F)] - 2|\mathrm{Cov}_d[\psi(F), (v\varepsilon/8) \cdot \psi'(F)]| \\
&= \mathrm{Var}_d[\psi'(F)] - \frac{v\varepsilon}{4} \cdot |\mathrm{Cov}_d[\psi(F), \psi'(F)]| \\
&\geq \mathrm{Var}_d[\psi'(F)] - \frac{v\varepsilon}{2} \\
&\geq \frac{v}{2},
\end{aligned}$$

where the last inequality follows from our definition of v. In particular, $\mathrm{Var}_d[\psi(F)] \geq \varepsilon^{3 \cdot 2^{|\mathcal{D}|} - 3}$ for sufficiently small ε. On the other hand, $|\mathbb{E}_d[\psi(F)]| \leq v\varepsilon/8 \leq (\varepsilon/4) \cdot \mathrm{Var}_d[\psi(F)]$, as desired. ◀

Proof of Theorem 4. If $(\mathcal{F}, \mathcal{D})$ satisfies HNK, then it also satisfies condition 2 of Proposition 12. By Proposition 7, $(\mathcal{F}, \mathcal{D})$ is extractable with error $O(\sqrt{\varepsilon}) + n/\varepsilon^{C+1}$. The forward direction follows by setting $n = \varepsilon^{C+1.5}$.

For the reverse direction, if $(\mathcal{F}, \mathcal{D})$ fails to satisfy HNK, by Proposition 12, then it also fails to satisfy MVR(ε) for some $\varepsilon > 0$. So by Proposition 10 it is not extractable. ◀

Alternatively, the reverse direction of Theorem 4 can be derived from Theorem 6 of [3] because if $(\mathcal{F}, \mathcal{D})$ fails (NHK) then it contains some $(\mathcal{F}', \mathcal{D}')$ which fails (NK).

3 Randomness-efficient extraction

In this Section we outline the proof of Theorem 5. The implication $1 \to 3$ in Theorem 5 is given by the following Proposition:

▶ **Proposition 13.** *For every $\varepsilon > 0$ and m, every GSV source that satisfies (NK$^+$) is extractable with error ε and output length m from $O((\log(1/\varepsilon) + m)/v^2)$ samples in time $\min\{O(m2^m \cdot n), n^{O(|\mathcal{F}|)}\}$, where v is the minimum of $\mathrm{Var}_d[\psi(F)]$ over all $d \in \mathcal{D}$.*

The case $m = 1$ is proved in Appendix A. The full proof is given in [1].

Implication $3 \to 2$ in Theorem 5 holds trivially. The remaining implication $2 \to 1$ follows readily from Propositions 15 and 16 below. These refer to an analytic condition that characterizes randomness-efficient extractability called the mean-variance divergence (MVD) condition, which can be viewed as the suitable analogue of the MVR condition from Section 2.

▶ **Definition 14.** A GSV source $(\mathcal{F}, \mathcal{D})$ satisfies the *Mean-Variance Divergence condition* $\mathrm{MVD}(\varepsilon, \delta)$ if there exists a function $\psi : \mathcal{F} \to [-1, 1]$ such that for every die $d \in \mathcal{D}$,

$$\left|\mathbb{E}_d[\psi(F)]\right| < \epsilon(\mathrm{Var}_d[\psi(F)] - \delta). \tag{MVD}$$

▶ **Proposition 15.** *If $(\mathcal{F}, \mathcal{D})$ fails (NK^+) then there exists a constant C such that for every $\varepsilon > 0$, $(\mathcal{F}, \mathcal{D})$ fails $\mathrm{MVD}(\varepsilon, C\varepsilon^2)$.*

▶ **Proposition 16.** *If $(\mathcal{F}, \mathcal{D})$ fails $\mathrm{MVD}(\varepsilon, \delta)$ then every extractor with error $\varepsilon/20$ for $(\mathcal{F}, \mathcal{D})$ requires $1/8\delta$ samples, assuming $\varepsilon > 0$ is sufficiently small.*

The proofs are given in Appendix B.

4 Open Questions

In this work, we completely classify GSV sources in terms of their extractability and give optimal deterministic extractors for GSV sources. We point out the following questions for further investigation:

- Let c be the smallest constant for which there exists a non-NK^+ source that is extractable from $O(1/\varepsilon^c)$ samples. Example E2 gives the upper bound $c \leq 7$.[5] Theorem 5 shows that $c \geq 2$. What is the value of c?
- The number of required samples in Theorem 4 is of the form $\varepsilon^{-O(2^{|\mathcal{D}|})}$, where $|\mathcal{D}|$ is the number of dice (see the proof of Proposition 12). Is this exponential dependence in $|\mathcal{D}|$ necessary?
- The multi-bit extractor in Theorem 5 runs in time $\min(nm2^m, n^{O(|\mathcal{F}|)})$. Can the dependence on the number of faces be improved, possibly by applying known seeded extraction algorithms?
- Proposition 7 states that sources satisfying condition $\mathrm{MVR}(\varepsilon)$ admit extraction with error $O(\sqrt{\varepsilon})$, while by Proposition 10 extraction error $\Omega(\varepsilon)$ is necessary. Can this quadratic gap be narrowed?
- It would be interesting to investigate extensions to infinite faces and/or dice.

References

1 Salman Beigi, Andrej Bogdanov, Omid Etesami, and Siyao Guo. Complete classification of generalized Santha-Vazirani sources. Technical Report TR17-136, Electronic Colloquium on Computational Complexity, 2017.

2 Salman Beigi, Omid Etesami, and Amin Gohari. Deterministic randomness extraction from generalized and distributed santha-vazirani sources. In *International Colloquium on Automata, Languages, and Programming*, pages 143–154. Springer, 2015.

[5] This source satisfies $\mathrm{MVR}(\varepsilon)$ with minimum variance ε^2 for every ε (with witness $\psi = (\varepsilon, -\varepsilon, 1, -1)$), so $O(1/\varepsilon^7)$ samples are sufficient for extraction error ε by Proposition 7.

3 Salman Beigi, Omid Etesami, and Amin Gohari. Deterministic randomness extraction from generalized and distributed santha-vazirani sources. *SIAM Journal on Computing*, 46(1):1–36, 2017.

4 Iddo Bentov, Ariel Gabizon, and David Zuckerman. Bitcoin beacon. *arXiv preprint arXiv:1605.04559*, 2016.

5 Jean Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(01):1–32, 2005.

6 Jean Bourgain. On the construction of affine extractors. *GAFA Geometric And Functional Analysis*, 17(1):33–57, 2007.

7 Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 670–683, 2016. doi:10.1145/2897518.2897528.

8 Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

9 Yevgeniy Dodis, Shien Jin Ong, Manoj Prabhakaran, and Amit Sahai. On the (im)possibility of cryptography with imperfect randomness. In *45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings*, pages 196–205, 2004. doi:10.1109/FOCS.2004.44.

10 Zeev Dvir. Extractors for varieties. *Computational complexity*, 21(4):515–572, 2012.

11 Zeev Dvir, Ariel Gabizon, and Avi Wigderson. Extractors and rank extractors for polynomial sources. *Computational Complexity*, 18(1):1–58, 2009.

12 Ariel Gabizon. Deterministic extractors for affine sources over large fields. In *Deterministic Extraction from Weak Random Sources*, pages 33–53. Springer, 2011.

13 Saeed Mahloujifar, Dimitrios I Diochnos, and Mohammad Mahmoody. Learning under p-tampering attacks. *arXiv preprint arXiv:1711.03707*, 2017.

14 Saeed Mahloujifar and Mohammad Mahmoody. Blockwise p-tampering attacks on cryptographic primitives, extractors, and learners. In *Theory of Cryptography Conference*, pages 245–279. Springer, 2017.

15 Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM J. Discrete Math.*, 13(1):2–24, 2000. doi:10.1137/S0895480197329508.

16 Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *J. Comput. Syst. Sci.*, 33(1):75–87, 1986. doi:10.1016/0022-0000(86)90044-9.

17 Salil P Vadhan. *Pseudorandomness*, volume 56. Now, 2012.

A An optimal bit extractor: Proof of Proposition 13 for $m = 1$

Define random variables $Z_0, \ldots, Z_n$ by $Z_0 = 0$ and $Z_t = Z_{t-1} + (\psi(F_t)/2) \cdot (1 - |Z_{t-1}|)$ where F_t $(1 \le t \le n)$ is the t-th output of the GSV source. The extractor outputs the sign of Z_n.

Under (NK$^+$) the sequence (Z_t) is still a martingale so that the expectation of Z_n is 0. But now the range of Z_t is restricted to the open interval $(-1, 1)$. To prove that the sign of Z_n has small bias, we begin by showing that the logarithm of $1/D_t$, on average, grows by a constant in every step where $D_t = 1 - |Z_t|$ is the distance between Z_t and its sign. Then we will use to argue the expectation of D_n is exponentially small. This fact together with $\mathbb{E}[Z_n] = 0$ allows us to conclude that the sign of Z_n is exponentially close to being unbiased.

▶ **Claim 17.** $\mathbb{E}[\ln(1/D_t) - \ln(1/D_{t-1}) \mid D_{<t}] \ge v/24$.

Proof. Using the identity $|x| = \text{sign}(x) \cdot x$, we upper bound D_t by

$$\begin{aligned} D_t &= 1 - |Z_t| \le 1 - \text{sign}(Z_{t-1}) \cdot Z_t \\ &= 1 - \text{sign}(Z_{t-1})\Big(Z_{t-1} + \frac{\psi(F_t)}{2} \cdot D_{t-1}\Big) = \Big(1 - \text{sign}(Z_{t-1}) \cdot \frac{\psi(F_t)}{2}\Big) \cdot D_{t-1}. \end{aligned} \tag{5}$$

Therefore,

$$\begin{aligned} \mathbb{E}\Big[\ln \frac{D_{t-1}}{D_t} \;\Big|\; D_{<t}\Big] &\ge \mathbb{E}\Big[-\ln\Big(1 - \text{sign}(Z_{t-1}) \cdot \frac{\psi(F_t)}{2}\Big) \;\Big|\; D_{<t}\Big] \\ &\ge \mathbb{E}\Big[\text{sign}(Z_{t-1}) \cdot \frac{\psi(F_t)}{2} + \frac{1}{6}\Big(\text{sign}(Z_{t-1}) \cdot \frac{\psi(F_t)}{2}\Big)^2 \;\Big|\; D_{<t}\Big] \\ &= \mathbb{E}\Big[\text{sign}(Z_{t-1}) \cdot \frac{\psi(F_t)}{2} \;\Big|\; D_1, \ldots, D_{t-1}\Big] + \frac{1}{6}\mathbb{E}\Big[\Big(\frac{\psi(F_t)}{2}\Big)^2 \;\Big|\; D_{<t}\Big] \\ &\ge 0 + v/24. \end{aligned}$$

The inequalities follow from the positivity of the (D_i)s, the identity $-\ln(1-x) \ge x + x^2/6$, the boundedness of ψ, and the NK$^+$ assumption, respectively. ◀

Let $X_t = \ln(1/D_t) - (vt/24)$. By Claim 17, $\mathbb{E}[X_t | D_{<t}] \ge X_{t-1}$ so that the sequence (X_t) is a sub-martingale with respect to (D_t). By (5) and the triangle inequality

$$|X_t - X_{t-1}| \le |\ln D_t/D_{t-1}| + v/24 \le \max\{|\ln 1/2|, |\ln 3/2|\} + v/24 = \ln 2 + v/24$$

By Azuma's inequality,

$$\Pr[D_n \ge e^{-vn/48}] = \Pr[X_n \le -vn/48] \le e^{-(vn/48)^2/2n(\ln 2 + v/24)^2} = 2^{-\Omega(v^2 n)}.$$

Finally by the triangle inequality and the law of conditional expectations,

$$|\mathbb{E}[\text{sign}(Z_n)]| \le \mathbb{E}[D_n] + |\mathbb{E}[Z_n]| \le 2^{-vn/48} + \Pr[D_n \ge e^{-vn/48}] + |\mathbb{E}[Z_n]| = 2^{-\Omega(v^2 n)}.$$

B A lower bound on the quality of extraction

The *kernel* of GSV source $(\mathcal{F}, \mathcal{D})$, denoted by $\text{Ker}\,\mathcal{D}$, is the set of all $\psi\colon \mathcal{F} \to \mathbb{R}$ such that $\mathbb{E}_d[\psi_d(F)] = 0$ for all dice $d \in \mathcal{D}$.

▶ **Proposition 18.** *A GSV source $(\mathcal{F}, \mathcal{D})$ satisfies (NK$^+$) if and only if for every die $d \in \mathcal{D}$ there exists a function $\psi_d \in \text{Ker}\,\mathcal{D}$ that is not constant on the support of d.*

Proof of Proposition 18. The forward direction follows by setting all ψ_d to equal the witness ψ for the (NK$^+$) condition. For the reverse direction, let $\psi = \sum_{d \in \mathcal{D}} N_d \psi_d$ where N_d are independent random variables, each uniformly distributed over some finite set $\mathcal{N} \subseteq \mathbb{R}$ of size more than $|D|$. By linearity, ψ is in $\text{Ker}\,\mathcal{D}$. Moreover, for each die d and each possible choice of the values $N_{d'}$ for $d' \ne d$, the sum $\sum N_d \psi_d$ can be constant on the support of d for at most one choice of N_d (for if two such choices existed then ψ_d itself must be constant on the support of d). Therefore, ψ is constant on d with probability at most $1/|\mathcal{N}|$. Since $|\mathcal{N}| > |\mathcal{D}|$, the existence of an (NK$^+$) witness ψ follows from the union bound. ◀

▶ **Claim 19.** *If (NK$^+$) fails for GSV source $(\mathcal{F}, \mathcal{D})$ then there exists a die $d \in \mathcal{D}$ such that for every pair of faces f^*, f_* in the support of d there exists a function $\beta\colon \mathcal{D} \to \mathbb{R}$ such that for all functions $\psi\colon \mathcal{F} \to \mathbb{R}$,*

$$\psi(f^*) - \psi(f_*) = \sum_{d' \in \mathcal{D}} \beta(d') \cdot \mathbb{E}_{d'}[\psi(F)]. \tag{6}$$

Proof. If $f^* = f_*$ the conclusion holds with $\beta = 0$. Otherwise, let $\mathcal{C}_d$ denote the linear space of functions that are constant on the support of die d. By Proposition 18, if (NK^+) fails then there exists a die d for which all functions $\psi \in \operatorname{Ker} \mathcal{D}$ also belong to $\mathcal{C}_d$, i.e., $\operatorname{Ker} \mathcal{D} \subseteq \mathcal{C}_d$. Then $\mathcal{C}_d^\perp \subseteq (\operatorname{Ker} \mathcal{D})^\perp$, where $\perp$ indicates the dual subspace. The space $(\operatorname{Ker} \mathcal{D})^\perp$ is the span of the probability mass functions pmf_d of all the dice. Therefore every $\phi \in \mathcal{C}_d^\perp$ can be written as a linear combination

$$\phi = \sum_{d \in \mathcal{D}} \beta(d) \cdot \mathrm{pmf}_d.$$

Then for every $\psi \colon \mathcal{F} \to \mathbb{R}$,

$$\sum_{f \in \mathcal{F}} \phi(f) \cdot \psi(f) = \sum_{d \in \mathcal{D}, f \in \mathcal{F}} \beta(d) \cdot \mathrm{pmf}_d(f) \cdot \psi(f) = \sum_{d \in \mathcal{D}} \beta(d) \cdot \mathbb{E}_d[\psi(F)].$$

The claim follows by specializing ϕ to the function that takes value 1 on f^*, -1 on f_*, and 0 elsewhere. This function is dual to $\mathcal{C}_d$. ◀

Proof of Proposition 15. Assume $(\mathcal{F}, \mathcal{D})$ fails (NK^+). Let d be the die stipulated by Claim 19 and C be the maximum of $(\sum_{d' \in \mathcal{D}} |\beta(d')|)^2$ over all pairs of faces f^*, f_* in the support of d.

Towards a contradiction suppose that $(\mathcal{F}, \mathcal{D})$ satisfies $\mathrm{MVD}(\varepsilon, \delta)$. Then the witness $\psi \colon \mathcal{F} \to [-1, 1]$ for $\mathrm{MVD}(\varepsilon, \delta)$ must satisfy the conditions $\mathrm{Var}_d[\psi(F)] > \delta$ and $|\mathbb{E}_{d'}[\psi(F)]| < \varepsilon \mathrm{Var}_{d'}[\psi(F)]$ for all dice $d' \in \mathcal{D}$. Let f^* and f_* be faces in the support of d that maximize and minimize the value of ψ, respectively. By Claim 19, relation (6) holds for some β that may depend on f^* and f_* but not on ψ. Then

$$\begin{aligned}\sqrt{\delta} < \sqrt{\mathrm{Var}_d[\psi(F)]} \le \psi(f^*) - \psi(f_*) &= \sum_{d' \in \mathcal{D}} \beta(d') \cdot \mathbb{E}_{d'}[\psi(F)] \\ &\le \sum_{d' \in \mathcal{D}} |\beta(d')| \cdot |\mathbb{E}_{d'}[\psi(F)]| < \sum_{d' \in \mathcal{D}} |\beta(d')| \cdot \varepsilon \mathrm{Var}_{d'}[\psi(F)] \le \sqrt{C}\varepsilon,\end{aligned}$$

where the last inequality follows from the definition of C and the boundedness of ψ. Therefore $\mathrm{MVD}(\varepsilon, \delta)$ fails for $\delta = C\varepsilon^2$. ◀

Proof of Proposition 16. The proof is a direct extension of the proof of Proposition 10. The main technical tool is the following claim:

▶ **Claim 20.** *For every extractor* $\mathrm{Ext} \colon \mathcal{F}^n \to \{0, 1\}$*, and every* $0 \le \alpha \le 1$*, if* $\mathbb{E}_{A_-}[\mathrm{Ext}] \ge \alpha$ *for every strategy* A_-*, then there exists a strategy* A_+ *for which*

$$\mathbb{E}_{A_+}[\mathrm{Ext}] \ge \alpha + \frac{\varepsilon}{1+\varepsilon} \cdot \big(\alpha(1-\alpha) - \delta n\big).$$

The proof of Claim 20 is a notationally intensive direct extension of the proof of Claim 11. We omit the details.

By Claim 20 it follows that for every $\varepsilon > 0$, if no strategy A_- has error less than $\alpha = 1/2 - \varepsilon/20$ against Ext then there exists a strategy A_+ with

$$\mathbb{E}_{A_+}[\mathrm{Ext}] \ge \frac{1}{2} - \frac{\varepsilon}{20} + \frac{\varepsilon}{1+\varepsilon} \cdot \Big(\frac{1 - \varepsilon^2/400}{4} - \frac{1}{8}\Big),$$

which is at least $1/2 + \varepsilon/20$ for sufficiently small ε. ◀