On Computationally Bounded Adverserial Capacity

Kyomin Jung
MIT
Email: kmjung@mit.edu

Devavrat Shah
MIT
Email: devavrat@mit.edu

***Abstract*— The main reason behind the complexity of coding-decoding scheme is the randomness in the source or channel noise. We model source or channel noise as pseudo-random to study the limitation of coding-decoding schemes in the presence of an adversary. In this adverserial model, we characterize the limitations of computationally bounded source and channel coding-decoding schemes in terms of classical Information theoretic quantities such as shannon capacity and entropy. It is well-known that from a small amount of truly random bits, a very large amount of pseudo-random bits can be generated (under certain hypothesis). Subsequently, we find that in our adverserial model with computationally bounded schemes, the channel capacity becomes arbitrarily smaller compared to the classical shannon capacity or compressible source becomes *incompressible*. As a byproduct, our results will lead to novel (negative) characterization of pseudo-random generators.**

I. Introduction

In this paper, we present a way to characterize the limitations of computationally bounded coding-decoding schemes for source compression and noisy channel in the presence of adversary.

A. Information Theory: Classical Result

The paper by Shannon [1] presented fundamental results characterizing the maximal compression of a random source as well as the maximal rate of transmission over a noisy channel. We present a quick summary of the known results for completeness. An interested reader can find details and other related results in books such as [2], [3].

Source coding theorem. The source coding theorem is about characterizing the minimal length required to described random sequences. In that sense, it also provides the characterization of "essential information in a random source". Next, we present the precise results.

Definition 1 (Entropy): Let X be a random variable taking values in finite set $\mathcal{X} = \{1, \ldots, \Sigma\}$. The entropy of (distribution of) X is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} \Pr(X = x) \log \Pr(X = x).$$

In this paper, we will restrict ourselves to finite $\mathcal{X}$. However, there are natural generalizations known for countably infinite and continuous $\mathcal{X}$.

Definition 2 (Lossy source coding): Let $\mathcal{X}^n$ denote the set of all strings of length $n \in \mathbb{N}$ with each element from $\mathcal{X}$. Given a distribution μ_n on $\mathcal{X}^n$ and $\epsilon > 0$, an ϵ-lossy source code is a mapping $\mathcal{C}_n : \mathcal{X}^n \to \mathcal{X}^* \cup \mathbf{e}$ such that (with respect to μ_n)

$$\Pr(\mathcal{E}) < \epsilon, \quad \text{and} \quad \mathcal{C}_n(x^n) \neq \mathcal{C}_n(\hat{x}^n), \ \forall x^n, \hat{x}^n \in \mathcal{E}^c,$$

where $\mathcal{E} = \{x^n \in \mathcal{X}^n : \mathcal{C}_n(x^n) = \mathbf{e}\}$. Here, $\mathbf{e}$ denote "loss" in coding and we denote $|\mathcal{C}_n(\mathbf{e})| = 0$.
The following theorem [1] characterizes the optimal source coding.

Theorem 3: Let $X_i, i \in \mathbb{N}$, be sequence of i.i.d. random variable distributed like a random variable X taking values in $\mathcal{X}$ and any $\epsilon > 0$. Then, for any ϵ-lossy source coding $\mathcal{C}_n$, there exists $n(\epsilon)$ such that for all $n \geq n(\epsilon)$,

$$\frac{1}{n}\mathbb{E}\left[|\mathcal{C}_n(X_1, \ldots, X_n)|\right] \geq H(X) - \epsilon.$$

Further, there exists $\hat{\mathcal{C}}_n$, such that for $n \geq n(\epsilon)$,

$$\frac{1}{n}\mathbb{E}\left[|\hat{\mathcal{C}}_n(X_1, \ldots, X_n)|\right] \leq H(X) + \epsilon.$$

Channel coding theorem. The channel coding theorem is about characterization of maximal rate at which data can be transmitted over a noisy channel. In this paper, we consider the case of discrete memory-less channel.

Definition 4 (Discrete memory-less channel): A discrete memory-less channel is characterized by the triple $(\mathcal{X}, \mathcal{Y}, \mathbb{Q})$, where $\mathcal{X}$ is the set of channel-input symbols, $\mathcal{Y}$ is set of channel-output symbols and $\mathbb{Q}$ is the conditional probability distribution of output y when x is transmitted over channel. We assume that channel is memory-less, that is

$$\Pr(y^n|x^n) = \prod_{i=1}^{n} \mathbb{Q}(y_i|x_i).$$

Definition 5 (Capacity): The capacity of a discrete memory-less channel $(\mathcal{X}, \mathcal{Y}, \mathbb{Q})$ is defined as

$$C(\mathbb{Q}) = \max_{\mu} H(Y) - H(Y|X),$$

where X is random variable distributed as μ over $\mathcal{X}$; Y be output random variable when X is transmitted as input over channel.

Definition 6 (Coding, decoding and probability of error): Given channel $(\mathcal{X}, \mathcal{Y}, \mathbb{Q})$, an (n, K) code $\mathcal{C}$ is one-to-one mapping from set of signals $\{1, \ldots, 2^K\}$ to 2^K distinct n-vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_{2^K}\}$ where $\mathbf{x}_i \in \mathcal{X}^n$. Rate of this code is defined as $R = K/n$. Encoder transmits $\mathbf{x}_i$ over channel to convey i to receiver. Let $\mathbf{Y}_i$ be random variable representing received output when $\mathbf{x}_i$ is transmitted over the channel.

A decoder, $\mathcal{D}$ is a mapping from received output, $\mathcal{Y}^n$ to $\{1,\ldots,2^K\} \cup \mathbf{e}$, where $\mathbf{e}$ indicates failure in decoding. The probability of error, P_e, is

$$P_e = 2^{-K} \sum_{1 \le i \le 2^K} \Pr(i \neq \mathcal{D}(\mathbf{Y}_i)).$$

The following theorem [1], [2], [3] characterizes the optimal transmission rate.

Theorem 7: Given channel $(\mathcal{X}, \mathcal{Y}, \mathbb{Q})$, an $\epsilon > 0$ and $R < C(\mathbb{Q})$ there exists an (n, K) coding scheme $\mathcal{C}$ with rate R and decoding scheme $\mathcal{D}$ with probability of error $P_e < \epsilon$. Further, for any $R > C(\mathbb{Q})$ there exists positive $\epsilon > 0$ such that for any (n, K) code with rate R there is no decoding scheme with $P_e < \epsilon$.

B. Pseudo Randomness

The notion of pseudo-randomness was first introduced by Blum-Micali [5] and Yao [4]. Intuitively, a distribution is called pseudo-random if it can not be distinguished from *true* distribution in a computationally efficient manner.

Definition 8 (Pseudo-random): Let $\mu_1 = (\mu_1^n)$ and $\mu_2 = (\mu_2^n)$, where μ_1^n, μ_2^n be distributions on $\mathcal{X}^n$. We say that μ_1 is poly-time indistinguishable from μ_2 if for all polynomial time statistical tests A and any polynomial p, $\exists n_0$ such that $\forall n \ge n_0$,

$$\left| \Pr_{x \in \mu_1^n}[A(x) = 1] - \Pr_{x \in \mu_2^n}[A(x) = 1] \right| < \frac{1}{p(n)}.$$

A distribution $\hat{\mu} = (\hat{\mu}^n)$ is called pseudo-random $\mu = (\mu^n)$, if $\hat{\mu}$ is poly-time indistinguishable from μ.
In the above definition of statical tests, it is assumed that they can be randomized.

Now, by definition a true distribution μ is pseudo-random μ as well. However, there are known constructions that generate large amount of pseudo-randomness from small amount of true randomness. First define formally what we mean by a pseudo-random generator. For this, we will assume that $\mathcal{X} = \{0, 1\}$. However, all definitions and constructions naturally extend for any finite $\mathcal{X}$.

Definition 9 (Pseudo-random generator): A polynomial time deterministic program $G : \{0,1\}^k \to \{0,1\}^{\hat{k}}$ is a pseudo-random generator (PSRG) with respect to distribution $\mu = (\mu^n)_{n \in \mathbb{N}}$ if the following holds: (1) $\hat{k} > k$, (2) $\hat{\mu}^{\hat{k}}$ is pseudo-random μ^k where $\hat{\mu}^{\hat{k}}$ is the distribution induced on $\mathcal{X}^{\hat{k}}$ when G is applied to x which is drawn according to uniform distribution on $\mathcal{X}^k$.

Many constructions of pseudo-random generators are known, conditional on *standard* computational hypothesis. Here, we present one such result. We refer interested reader to book by Goldwasser and Bellare [8] for any missing details in this paper as well as importance of pseudo-randomenss in the context of cryptography and computational complexity.

Theorem 10: Let there exists a length preserving one-way permutation $f = (f^n)$, where $f^n : \{0,1\}^n \to \{0,1\}^n$. Then, for every polynomial P, using f it is possible to construct a PSRG $G : \{0,1\}^k \to \{0,1\}^{P(k)}$ such that $G(x)$ is pseudo-random uniform on $\{0,1\}^{P(k)}$ when x is distributed as uniform on $\{0,1\}^k$.

Theorem 10 suggests that if we are given n truly random bits, then they can be converted to n^α pseudo-random bits for any $\alpha \in \mathbb{N}$. We will use this to derive interesting implications of our results.

Note. A seminal paper by Nisan and Widgerson [6] led to many interesting results in the context of pseudo-randomness. We also note that a survey paper by Shaltiel [7] provides recent developments in the context of constructing *extractors* (intuitively, algorithm extracting almost uniform randomness out of non-uniform distribution), using pseudo-random generators.

C. Our Results

The results of this paper is about characterizing the limitations of computationally bounded source and channel coding schemes. Now, classical computational complexity is defined with respect to the worst-case or in an adverserial setup. Hence, to understand the computational limitations of coding schemes, we need to consider an appropriate adverserial setup. Now, the questions remains how should one define adversary for coding schemes.

As noted earlier, the main reason behind coding-decoding complexity is the randomness in source (in the context of source coding) and randomness in the channel noise (in the context of channel coding). For example, if there was no noise, then one does not need to code (or decode). On the contrary, even if there is (almost) deterministic noise, if it can not be *learnt* efficiently by coder-decoder, then it may require complex coding-decoding algorithms.

The above considerations motivate us to model the randomness in source and channel noise as pseudo-random to understand the limitations of computationally bounded source coding and channel coding schemes.

Source coding theorem. We first define a polynomial time lossy compression scheme.

Definition 11 (Polynomial-time lossy compression): Given an $\epsilon > 0$, an ϵ-lossy compression scheme $\mathcal{C}_n : \mathcal{X}^n \to \mathcal{X}^* \cup \mathbf{e}$, $n \in \mathbb{N}$, for a source generating elements from $\mathcal{X}$, is called polynomial time if the operations done by $\mathcal{C}_n$ to map $x^n \in \mathcal{X}^n$ to $\mathcal{C}_n(x^n)$ is polynomial in n. Such a compression scheme is oblivious of the source distribution, however its allowed to sample source distribution polynomial in n times to possibly determine $\mathcal{C}_n$.
It should be clear that a polynomial-time lossy compression scheme $\mathcal{C}_n$ will always map $x^n \in \mathcal{X}^n$ to $y^m \in \mathcal{X}^m$, where $m = poly(n)$.

Theorem 12: Let an adverserial source generates strings $\hat{X} = (\hat{X}^n)_{n \in \mathbb{N}}$. Let the distribution of $\hat{X}^n$ be $\hat{\mu}^n$ which is pseudo-random μ^n. Let $X = (X^n)_{n \in \mathbb{N}}$ be strings where X^n is generated according to true distribution μ^n. Let $\epsilon > 0$ be given. Then, there exists large enough $n_0(\epsilon)$ such that for all $n \ge n_0(\epsilon)$ and any polynomial-time ϵ-lossy compression scheme, $\mathcal{C}_n$,

$$\left| \mathbb{E}[|\mathcal{C}_n(\hat{X}^n)|] - \mathbb{E}[|\mathcal{C}_n(X^n)|] \right| < \epsilon.$$

Channel coding theorem. We will consider binary channel with additive noise. In such channel, input and output alphabets are binary, i.e. $\mathcal{X} = \mathcal{Y} = \{0,1\}$. The noise is additive, that is when $\mathbf{x} \in \{0,1\}^n$ is transmitted, noise $\mathbf{t} \in \{0,1\}^n$ (independent of $\mathbf{x}$) is added to produce output $\mathbf{y} \in \{0,1\}^n$. Thus, $\mathbf{y} = \mathbf{t} \oplus \mathbf{x}$, where $\oplus$ denotes component-wise addition over $\mathbb{Z}_2$. In such channel, the additive noise characterizes the channel capacity. Hence, we model adverserial setup by modeling the additive noise as pseudo-random. Next, we define polynomial-time coding and decoding schemes for binary channel.

Definition 13 (Polynomial-time coding-decoding): An (n, K) code $\mathcal{C}$ and corresponding decoder $\mathcal{D}$ are called polynomial-time coding-decoding if the following holds: mapping any $i \in \{1, \ldots, 2^K\}$ to $\mathcal{C}(i)$ and mapping any output $\mathbf{Y}_i$ to $\mathcal{D}(\mathbf{Y}_i)$ requires operations that are polynomial in $\max(n, K)$. The rate of such a code is K/n as defined earlier.

Now, we state our result that relates the capacity of polynomial-time coding-decoding schemes to the shannon capacity.

Theorem 14: Consider an additive binary channel. Let $\hat{T}^n \in \{0,1\}^n$ denote the additive noise random variable. Let the distribution of $\hat{T}^n$ be pseudo-random $B(n,p)$, the binomial with parameter $p \in (0,1)$. Let $\epsilon > 0$ be given. Consider any polynomial-time coding decoding scheme $\mathcal{C}$ and $\mathcal{D}$ with an (n, K) code. Let $P_e(Z^n)$ denote probability of error when noise is Z^n. Then, there exists large enough $n_0(\epsilon)$ such that for $n \geq n_0(\epsilon)$,

$$\left| P_e(\hat{T}^n) - P_e(T^n) \right| < \epsilon,$$

where T^n is distributed as $B(n,p)$.

A straight-forward corollary of Theorem 14 is as follows:

Corollary 15: Consider an additive binary channel. Let $\hat{T}^n \in \{0,1\}^n$ denote the additive noise random variable. Let the distribution of $\hat{T}^n$, $\hat{\mu}^n$, be pseudo-random $B(n,p)$, binomial with parameter $p \in (0,1)$. Then, for $R > C(p) = 1 - H(p)$, there exists an $\epsilon > 0$ and $n_0(\epsilon)$ such that for all $n \geq n_0(\epsilon)$ there is no ϵ-good polynomial-time coding decoding scheme that operates at rate R.

D. Organization

Rest of the paper is organized as follows. Section II presents proofs of the main theorems. Section III presents implications of our results, discussion and directions for future work.

II. Proofs of Theorems

In this section, we present proofs of Theorems 12 and 14.

A. Proof of Theorem 12

The proof follows by a straightforward use of the definition of pseudo-randomness. To this end, consider a polynomial time ϵ-lossy compression scheme $\mathcal{C}_n : \mathcal{X}^n \to \mathcal{X}^{\leq m} \cup \mathbf{e}$, where $m = p(n)$ some polynomial in n and

$$\mathcal{X}^{\leq m} = \cup_{k \leq m} \mathcal{X}^k.$$

By definition, $|\mathcal{C}_n(x)| \leq p(n)$ for all $x \in \mathcal{X}^n$. Now, define a randomized polynomial time statistical test $\mathcal{A}(x)$ as follows: *choose $i \in \{1, \ldots, p(n)\}$ uniformly at random. Declare $\mathcal{A}(x) = 1$ if $|\mathcal{C}_n(x)| \geq i$ and 0 otherwise.*

Note that $\mathcal{A}$ is a randomized (with $O(\log n)$ bits of randomness) polynomial time algorithm because $\mathcal{C}_n$ is polynomial time. Hence, the definition of pseudo-randomness implies that for given $\epsilon > 0$, there exists a $n_0(\epsilon)$ (independent of $\mathcal{A}, \mathcal{C}_n$) such that for $n \geq n_0(\epsilon)$,

$$\left| \Pr[\mathcal{A}(\hat{X}^n) = 1] - \Pr[\mathcal{A}(X^n) = 1] \right| \quad < \quad \frac{\epsilon}{n^2 p(n)}, \quad (1)$$

where we use the fact that distribution of $\hat{X}^n$ is pseudo-random μ^n, the distribution of X^n. Now, from (1) we obtain

$$\left| \sum_{i=1}^{p(n)} \frac{\Pr[|\mathcal{C}_n(\hat{X}^n)| \geq i]}{p(n)} - \sum_{i=1}^{n} \frac{\Pr[|\mathcal{C}_n(X^n)| \geq i]}{p(n)} \right| \quad < \quad \frac{\epsilon}{n^2 p(n)} \quad (2)$$

Since $|\mathcal{C}_n(\cdot)| \leq n$, we obtain that

$$\mathbb{E}[|\mathcal{C}_n(Z)|] \quad = \quad \sum_{i=1}^{p(n)} \Pr[|\mathcal{C}_n(Z)| \geq i], \quad (3)$$

for $Z = \hat{X}^n, X^n$. From (2) and (3), we immediately obtain that

$$\left| \mathbb{E}[|\mathcal{C}_n(\hat{X}^n)|] - \mathbb{E}[|\mathcal{C}_n(X^n)|] \right| \quad < \quad \frac{\epsilon}{n^2} \quad \leq \quad \epsilon. \quad (4)$$

This completes the proof of Theorem 12.

B. Proof of Theorem 14

We will prove Theorem 14 just like Theorem 12. Let $\epsilon > 0$ be given. Let there be a polynomial-time coding and decoding schemes $\mathcal{C}$ and $\mathcal{D}$, where $\mathcal{C}$ is an (n, K) code with rate $R = K/n$.

Define an algorithm $\mathcal{A}_i, 1 \leq i \leq 2^K$ that takes as input the noise sequence Z^n and outputs 0 or 1 as follows: *Map i to $\mathbf{x}_i = \mathcal{C}(i)$ using the code and transmit over channel with additive noise Z^n. The corresponding output is $\mathbf{Y}_i = \mathbf{x}_i \oplus Z^n$. Decode $\mathbf{Y}_i$ by mapping it via decoder $\mathcal{D}(\mathbf{Y}_i) \in \{1, \ldots, 2^K\}$. Declare $\mathcal{A}(Z^n) = 1$ if $\mathcal{D}(\mathbf{Y}_i) = i$ and 0 otherwise.*

From above description, the $\mathcal{A}_i$ outputs 0 if and only if there is an error in decoding. The algorithm $\mathcal{A}_i$ is polynomial time as $\mathcal{C}, \mathcal{D}$ are polynomial time. Given $\epsilon > 0$, by definition of pseudo-randomness there exists an $n_0(\epsilon)$ such that for all $n \geq n_0(\epsilon)$ for all $i, 1 \leq i \leq 2^K$, the following holds:

$$\left| \Pr(\mathcal{A}_i(\hat{T}^n) = 1) - \Pr(\mathcal{A}_i(T^n) = 1) \right| \quad < \quad \epsilon. \quad (5)$$

Equivalently,

$$\left| \Pr(\mathcal{A}_i(\hat{T}^n) = 0) - \Pr(\mathcal{A}_i(T^n) = 0) \right| \quad < \quad \epsilon. \quad (6)$$

That is, for all $1 \leq i \leq 2^K$,

$$\left| \Pr_{\hat{T}^n}(\mathcal{D}(\mathbf{Y}_i) \neq i) - \Pr_{T^n}(\mathcal{D}(\mathbf{Y}_i) \neq i) \right| \quad < \quad \epsilon, \quad (7)$$

where $\Pr_Z$ means probability induced by distribution of Z. From (7) and definition of probability error, $P_e(\cdot)$ it immediately follows that

$$\left| P_e(\hat{T}^n) - P_e(T^n) \right| \quad < \quad \epsilon. \tag{8}$$

This completes the proof of Theorem 14.

III. Discussion

In this section, we present interpretation of Theorems 12-14 and related discussion. The results, as discussed below will possibly lead to interesting characterization of pseudo-random generators.

A. Source Coding Theorem

First, we consider Theorem 12. Consider a pseudo-random source that generates binary symbols. Let $X^n = (X_1, \ldots, X_n)$ be sequence of symbols generated by source for $n \in \mathbb{N}$. Let the distribution of X^n be pseudo-random $B(n, 1/2)$. Let $\epsilon > 0$ be given. Consider an ϵ-lossy polynomial time coding (compression) scheme and let $\ell(X^n)$ denote the length of the coded sequence. Then, by Theorem 12 and Theorem 3,

$$\frac{\mathbb{E}[\ell(X^n)]}{n} \quad \geq \quad 1 - \epsilon. \tag{9}$$

By Theorem 10, it is possible to generate pseudo-random $B(n, 1/2)$ from $B(n^\alpha, 1/2)$ for arbitrarily small $\alpha > 0$. But the essential information in $B(n^\alpha, 1/2)$ is only n^α bits. Thus (9) implies under the adverserial model of this paper, the compression can become arbitrary *bad*. Now, lets contrast this with well-known Lempel-Ziv coding theorem (from Chapter 12, [2]).

Theorem 16: Let $\{X_n\}, n \in \mathbb{N}$, be a binary stationary ergodic process. Then, there exists a coding scheme that codes $(X_1, \ldots, X_n)$ in $O(n^2)$ operations. Let $\ell(X_1, \ldots, X_n)$ be the length of the coded sequence. Then,

$$\limsup_{n \to \infty} \frac{\ell(X_1, \ldots, X_n)}{n} \leq H(X), \text{ with prob. } 1,$$

where $H(X) = \lim_{n \to \infty} H(X_n | X_1, \ldots, X_{n-1})$.
Our implication and Theorem 16 suggests that the pseudo-random sequences can not form stationary ergodic process ! Further implications of such well-known results in characterizing structural properties of pseudo-random generators remains the topic of future research.

B. Channel Coding Theorem

Here, we consider Theorem 14. To this end, suppose we are given additive binary channel with noise sequence $\hat{T}^n$ be pseudo-random $B(n, 1/2)$. Then, by Theorem 7 and Theorem 14, for any $\epsilon > 0$ there is no (n, K) code with $n \geq n_0(\epsilon)$ that transmits at rate larger than ϵ. By Theorem 10, such a noise can generated from $B(n^\alpha, 1/2)$ for any $\alpha > 0$. Suppose, originally we were given a channel with noise distributed as $B(n, p)$ for a very small $p > 0$. The capacity of channel with noise as $B(n, p)$ is $\approx 1 - \epsilon$ for p small enough. It is well-known (e.g. result of von Neumann [9]) that such a noise can be converted into equivalent $B(m, 1/2)$ so that $H(B(m, 1/2)) \approx H(B(n, p))$. For any positive $p > 0$, $m = n^\beta$ where $\beta > 0$ a small constant. This can be used by PSRG as explained above to create pseudo-random noise.

Putting the above together, we find that when noise is generated adverserialy, then the capacity becomes arbitrarily small !

Such a result can provide plausible explanation to the following scenario (suggested by R. Koetter): suppose you are a painter and want to post your picture on-line. However, you are afraid of *thieves* who will steal your picture and claim its ownership. To avoid this, you would do *digital watermarking* by adding some *noise* to the picture. If a lot of noise is added then you may loose the quality of picture and if little noise is added then you may loose *security*. The above discussion suggests that by adding pseudo-random noise, you may preserve both quality of picture and security.

Acknowledgment

We thank Ralf Koetter and Sanjoy Mitter for stimulating discussions after the seminar by Avi Widgerson, which eventually led to this work.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623–656, 1948.

[2] T. Cover and J. Thomas, "Information Theory," *Wiley-Interscience*, 1991.

[3] R. Gallager, "Information Theory and Reliable Communication," *John Wiley and Sons*, 1971.

[4] A. C. Yao, "Theory and application of trapdoor functions (extended abstract)," *In proceedings of FOCS,* 1982.

[5] M. Blum and S. Micali, "How to generate cryptographically strong sequences of pseudo-random bits," *SIAM Journal on Computing*, 13(4), 1984.

[6] N. Nisan and A. Wigderson, "Hardness vs randomness," *Journal of Computer and System Sciences*, 49(2), 1994.

[7] R. Shaltiel, "Recent developments in explicit constructions of extractors," *Bulletin of the European Association for Theoretical Computer Science*, 77, 2002.

[8] S. Goldwasser and M. Bellare, "Lecture Notes on Cryptography", *Available on-line at* `www.cs.ucsd.edu/users/mihir/papers/gb.pdf` .

[9] J. von Neumann, "Various techniques used in connection with random digits," *National Bureau of Standards Applied Math. Series*, 12, 1951.