[25] I. Csiszár, "Maxent, mathematics, and information theory," in *Maximum Entropy and Bayesian Methods: Proceedings of the 15th International Worksop (Santa Fe, NM)*, K. M. Hanson and R. N. Silver, Eds. Boston, MA: Kluwer Academic, 1996, pp. 35–50. 1995.

[26] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.

[27] C. D. Aliprantis and O. Burkinshaw, *Principles of Real Analysis*. San Diego, CA: Academic, 1990.

[28] S. Kullback, *Information Theory and Statistics*, 2nd ed. Mineola, NY: Dover, 1968.

[29] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden-Day, 1964.

[30] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[31] R. B. Ash, *Information Theory*. New York, NY: Dover, 1990.

On the Jensen–Shannon Divergence and Variational Distance

Shi-Chun Tsai, Wen-Guey Tzeng, and Hsin-Lung Wu

***Abstract*—We study the distance measures between two probability distributions via two different distance metrics, a new metric induced from Jensen–Shannon divergence, and the well known L_1 metric. We show that several important results and constructions in computational complexity under the L_1 metric carry over to the new metric, such as Yao's next-bit predictor, the existence of extractors, the leftover hash lemma, and the construction of expander graph based extractor. Finally, we show that the useful parity lemma in studying pseudorandomness does not hold in the new metric.**

***Index Terms*—Jensen–Shannon divergence, expander, extractors, leftover hash lemma, parity lemma.**

I. Introduction

For any two distributions P and Q over the sample space $\{\omega_1, \cdots, \omega_n\}$, the variational distance (under L_1 metric) between P and Q denoted by $SD(P,Q)$ is defined as

$$\frac{1}{2}\sum_{i=1}^{n} |\Pr[P = \omega_i] - \Pr[Q = \omega_i]|.$$

This definition is equivalent to the existence of the best distinguisher B such that $B(\omega_i) = 1$ if and only if $\Pr[P = \omega_i] \geq \Pr[Q = \omega_i]$ and

$$|\Pr_{\omega_i \leftarrow P}[B(\omega_i) = 1] - \Pr_{\omega_i \leftarrow Q}[B(\omega_i) = 1]| = SD(P,Q).$$

We say that two distributions P and Q on a sample space are ϵ-close in L_1-norm if $SD(P,Q) \leq \epsilon$. In computational complexity, many results have been obtained based on the L_1 metric, such as pseudorandomness and extractors [11] and Yao's next-bit predictor [13], etc. It prompts a natural question: Why we should use the L_1 metric in the first place. Can we use another metric of distributions instead of the variational distance? Suppose we have a new distance metric for probability distributions. Do the computational complexity results still hold under the new distance metric? Endres and Schindelin recently proposed a new metric ND for probability distributions [4]. The square of the new distance measure is the so-called Jensen–Shannon divergence. This motivates us to answer the above question for this new metric.

Jensen–Shannon divergence was proposed by Lin [6] for breaking the condition of absolute continuity of Kullback divergence. This research is information-theoretic. For application in computational complexity, especially communication complexity, it is natural to employ Jensen–Shannon divergence in the related problems. This is because Jensen–Shannon divergence captures some properties of the mutual information. For an application to communication complexity, we refer the readers to the paper of Bar-Yossef [2] in which he used a technique based on Jensen–Shannon divergence to prove lower bounds on the query complexity of sampling algorithms [2].

In this correspondence, we study the metric based on Jensen–Shannon divergence and use it to investigate some randomized computational complexity issues. We show that several important results and constructions in computational complexity under the L_1 metric carry over to the new metric, such as Yao's next-bit predictor [13], the existence of extractors [11], the leftover hash lemma [9], and the construction of expander graph based extractors. Finally, we show that the useful parity lemma [12] in studying pseudorandomness does not hold in the new metric.

II. Notation and Preliminary Results

Here we focus on discrete distributions whose sample space is finite. We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. The base of log function is 2. In this correspondence, for every positive integer m, the notation U_m always denotes the uniform distribution over $\{0,1\}^m$. We say a distribution D_n in $\{0,1\}^n$ is a k-source if for all $x \in \{0,1\}^n$, $D_n(x) \leq 2^{-k}$. The notation $\|\cdot\|$ always means the ℓ_2 norm. The following fact is useful in this correspondence.

Fact 1: $\ln 2 = \sum_{j=1}^{\infty} \frac{1}{2j(2j-1)}$. For any distribution P with sample space $\Omega_n = \{\omega_1, \ldots, \omega_n\}$, define the entropy of P to be

$$H(P) = \sum_{i=1}^{n} -\Pr[X = \omega_i] \log \Pr[X = \omega_i]$$

where we define $p \log p = 0$ if $p = 0$. For the properties of entropy function H, we refer readers to the textbooks by Cover and Thomas [3] and Yeung [14].

Definition 1: Let P and Q be two distributions with the same probability space. The quantity

$$\left(H\left(\frac{P+Q}{2}\right) - (H(P) + H(Q))/2\right)$$

is the Jensen–Shannon divergence. ND is defined as

$$ND(P,Q) = \sqrt{H\left(\frac{P+Q}{2}\right) - \frac{H(P)+H(Q)}{2}}.$$

Endres and Schindelin proved that ND is a metric [4]. Topsøe gave a lemma to characterize ND [10]. For convenience, every distribution in finite sample space can be viewed as a vector. So we write distributions P and Q as $P = \langle p_1, \ldots, p_n \rangle$ and $Q = \langle q_1, \ldots, q_n \rangle$ where $p_i = \Pr[P = \omega_i]$ and $q_i = \Pr[Q = \omega_i]$ for $1 \leq i \leq n$.

Lemma 1: [10] For any distributions P and Q over Ω_n

$$\frac{2}{\log e}(ND(P,Q))^2 = \sum_{j=1}^{\infty} \frac{1}{2j(2j-1)} \left(\sum_{i=1}^{n} \frac{|p_i - q_i|^{2j}}{(p_i + q_i)^{2j-1}} \right).$$

Manuscript received July 27, 2004; revised January 24, 2005. The work of S.-C. Tsai and H.-L. Wu was supported in part by the National Science Council of Taiwan under Contract NSC-93-2213-E-009-035. The work of W.-G. Tzeng was supported in part by the National Science Council of Taiwan under Contract NSC-93-2213-E-009-008 and a MediaTek Grant, 2003. The material in this correspondence was presented in part at the ICS Conference 2000, Taipei, Taiwan, R.O.C., November 2004.

The authors are with the Department of Computer Science, National Chiao-Tung University, Hsinchu 30050, Taiwan, R.O.C. (e-mail: sctsai@csie.nctu.edu.tw; tzeng@cis.nctu.edu.tw; hsinlung@csie.nctu.edu.tw).

Communicated by R. W. Yeung, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2005.853308

0018-9448/$20.00 © 2005 IEEE

The following theorem characterizes the relation between ND and SD.

Theorem 1: [10]

$$\sqrt{SD(P,Q)} \geq ND(P,Q) \\ \geq \sqrt{\tfrac{(1+SD(P,Q))\log(1+SD(P,Q))+(1-SD(P,Q))\log(1-\mathrm{SD}(P,Q))}{2}}.$$

Actually, the above bounds are tight. For the left-hand side inequality, we consider the following two distributions:

$$P = \left\langle \epsilon, \underbrace{\frac{1-\epsilon}{n-2}, \dots, \frac{1-\epsilon}{n-2}}_{n-2}, 0 \right\rangle$$

and

$$Q = \left\langle 0, \underbrace{\frac{1-\epsilon}{n-2}, \dots, \frac{1-\epsilon}{n-2}}_{n-2}, \epsilon \right\rangle.$$

Clearly, $SD(P,Q) = \epsilon$. We can compute $ND(P,Q) = \sqrt{\epsilon}$. Hence, the left-hand side is tight.

For the right-hand side we set

$$P = \left\langle \underbrace{\frac{1+\epsilon}{2n}, \dots, \frac{1+\epsilon}{2n}}_{n}, \underbrace{\frac{1-\epsilon}{2n}, \dots, \frac{1-\epsilon}{2n}}_{n} \right\rangle$$

and

$$Q = \left\langle \underbrace{\frac{1-\epsilon}{2n}, \dots, \frac{1-\epsilon}{2n}}_{n}, \underbrace{\frac{1+\epsilon}{2n}, \dots, \frac{1+\epsilon}{2n}}_{n} \right\rangle.$$

Clearly, $SD(P,Q) = \epsilon$, and we have

$$ND(P,Q) = \sqrt{\frac{(1+\epsilon)\log(1+\epsilon) + (1-\epsilon)\log(1-\epsilon)}{2}}.$$

Therefore, the right-hand side is a tight bound.

III. Randomized Computation Via ND

Randomized computation has been a very useful method for algorithm design. Randomized algorithms are the only known efficient methods for many difficult problems [7]. In this section, we illustrate that several important results in randomized computation based on SD carry over to ND. We also show a nonapplicable case.

A. Distinguisher Versus Predictor

Yao [13] proved that a Boolean function G is a good distinguisher between two distributions (where one of which is uniform) if and only if G is a good next-bit predictor. First of all we give some definitions.

Definition 2: For any distribution D_n on the probability space $\{0,1\}^n$, an ϵ-good distinguisher between D_n and U_n is a Boolean function C such that

$$\left| \Pr_{x \leftarrow D_n}[C(x)=1] - \Pr_{x \leftarrow U_n}[C(x)=1] \right| \geq \epsilon.$$

Definition 3: For any distribution D_n, an ϵ-good next-bit predictor for D_n is a function, for some $i \in [n]$ and given the first $(i-1)$ bits of the input, such that

$$| \Pr_{x \leftarrow D_n}[G(x_1, \dots, x_{i-1}) = x_i]| \geq \epsilon.$$

With a distinguisher as an oracle, Yao proved the following lemma.

Lemma 2: [13] If C is an ϵ-good distinguisher between D_n and U_n, then there exists an (ϵ/n)-good next-bit predictor for D_n.

By Theorem 1, we have the following result.

Theorem 2: Suppose $ND(D_n, U_n) \geq \epsilon$. Then we have a next-bit predictor G with the following property: there exists $i \in [n]$ such that

$$\Pr[G(x_1, \dots, x_{i-1}) = x_i] \geq (\epsilon^2/n)$$

where $x_1, \dots, x_i$ are sampled from D_n.

Proof: By Theorem 1, we have

$$SD(D_n, U_n) \geq ND(D_n, U_n)^2 \geq \epsilon^2.$$

By Lemma 2, there exists an $((\epsilon^2)/n)$-good next-bit predictor G for D_n. □

B. Extractors

We continue to show the existence of extractors under the setting of ND with some appropriate parameters. Similar to the definition of extractor [8], we have the following definition.

Definition 4: $\mathrm{EXT} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is called a (k,ϵ)-extractor for ND if for every k-source D_n

$$ND(\mathrm{EXT}(D_n, U_t), U_m) \leq \epsilon.$$

For ND we have the following analogous result.

Proposition 1: For every $n, \epsilon > 0$ and $k \leq n$, there exists a (k,ϵ)-extractor

$$\mathrm{EXT} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$$

for ND with $t = \log n - k - 4\log\epsilon + O(1)$ and $m = k + t + 4\log\epsilon - O(1)$.

Proof: We prove the proposition by the probabilistic method [1], [7]. Consider the random extractor f which maps $x \in \{0,1\}^{n+t}$ into $\{0,1\}^m$ randomly and independently. Since a k-source can be represented as a convex combination of flat k-sources and ND is a metric, it is sufficient to prove the proposition for flat sources. For any distribution P in $\{0,1\}^m$ and any Boolean function $T : \{0,1\}^m \to \{0,1\}$ we denote P_T as a distribution in $\{0,1\}$ with

$$\Pr[P_T = 1] = \sum_{x : T(x)=1} P(x).$$

We first prove the following claim.

Claim 1: For any flat $(k+t)$-source Q, if m and t satisfy the conditions of Proposition 1, then

$$\Pr[ND(f(Q), U_m) > \epsilon] < 2^{2^m} \cdot 2^{-\Omega(2^{k+t} \cdot \epsilon^4)}.$$

Proof: Let the support of distribution Q be

$$\mathrm{Supp}(Q) = \{x : Q(x) > 0\}.$$

For each $x \in \mathrm{Supp}(Q)$, the distribution of $f(x)$ is the same as U_m. Also, $\{f(x) : x \in \mathrm{Supp}(Q)\}$ is a set of random variables which are independent and identically distributed (i.i.d.). For each Boolean function $T : \{0,1\}^m \to \{0,1\}\{T(f(x)) : x \in \mathrm{Supp}(Q)\}$ is also a set of 0–1 random variables which are i.i.d. and

$$\mathrm{Exp}[T(f(x))] = \frac{|\{z : T(z) = 1\}|}{2^m} = \Pr[(U_m)_T = 1].$$

By the Chernoff bound [1], [7]

$$\Pr\left[\left|\frac{\sum_{x\in \mathrm{Supp}(Q)} T(f(x))}{2^{k+t}} - \frac{|\{z : T(z)=1\}|}{2^m}\right| > \epsilon^2\right] < 2^{-\Omega(2^{k+t}\epsilon^4)}.$$

By Theorem 1, we can get

$$\begin{aligned}\Pr[ND(f(Q),U_m) > \epsilon] &\le \Pr[\mathrm{SD}(f(Q),U_m)>\epsilon^2]\\ &\le \Pr[\exists T, \mathrm{SD}(f(Q)_T,(U_m)_T) > \epsilon^2]\\ &< 2^{2^m}\cdot 2^{-\Omega(2^{k+t}\cdot\epsilon^4)}.\end{aligned}$$

□

The probability that f is not a good extractor for some flat k-source is at most

$$\begin{aligned}\binom{2^n}{2^k}\cdot 2^{2^m}\cdot 2^{-\Omega(2^{k+t}\cdot\epsilon^4)} &\le \binom{2^n}{2^k}2^{-\Omega(2^{k+t}\cdot\epsilon^4)}\\ &\le \frac{2^{n\,2^k}}{2^k}\,2^{-\Omega(2^{k+t}\cdot\epsilon^4)}\\ &< 1\end{aligned}$$

This proves the existence of the extractor for ND. □

The crucial part of the proof is the inequality between SD and ND. Then we can use the property of SD to show the existence of extractor with good parameters. There seems no constructive proof on the existence of the extractor for ND.

C. Leftover Hash Lemma

Linearity plays an important role in the proof of the Leftover Hash Lemma and expander-based extractors. It seems that ND does not have such linear property. However, in some setting ND has a good upper bound in terms of the ℓ_2 norm. This bound can help us prove some results about extractors for ND.

Definition 5: [5] $\mathcal{H} = \{h : \mathcal{D} \to \mathcal{R}\}$ is universal family of hash functions if, for every $x, y \in \mathcal{D}$, $x \neq y$

$$\Pr_{h\leftarrow\mathcal{H}}[h(x)=h(y)] = \frac{1}{|\mathcal{R}|}.$$

$\mathcal{H}$ is almost universal if

$$\Pr_{h\leftarrow\mathcal{H}}[h(x)=h(y)] \le \frac{1}{|\mathcal{R}|} + \frac{1}{|\mathcal{D}|}.$$

Now let $\mathcal{D} = \{0,1\}^n, \mathcal{R} = \{0,1\}^m$ and $|\mathcal{H}| = 2^t$. The Leftover Hash Lemma states the following.

Theorem 3: [5] Suppose $\mathcal{H}$ is almost universal, X is a flat k-source on $\{0,1\}^n$, and $\boldsymbol{h}$ is a random function drawn from $\mathcal{H}$. Then

$$SD((\boldsymbol{h},\boldsymbol{h}(X)),U_{t+m}) \le \frac{1}{2^{(k-m)/2}}.$$

Define

$$\mathrm{Col}[(\boldsymbol{h},\boldsymbol{h}(X))] = \Pr[(\boldsymbol{h},\boldsymbol{h}(X)) = (\boldsymbol{h}',\boldsymbol{h}'(X'))]$$

where $\boldsymbol{h}', X'$ are i.i.d. to $\boldsymbol{h}, X$, respectively. The crucial part of the proof of Theorem 3 is to show the following lemma.

Lemma 3: [5] $\mathrm{Col}[(\boldsymbol{h},\boldsymbol{h}(X))] \le (1+2^{(1+m-k)})/(2^{t+m})$.

Define $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^{t+m}$ by $\mathrm{Ext}(x,h) = (h,h(x))$. We show that Ext is an extractor for ND. Here, instead of directly applying the inequality between ND and SD, we establish the relation between ND and the ℓ_2-norm.

Theorem 4: Suppose $\mathcal{H}$ is an almost universal family of hash functions from $\{0,1\}^n$ to $\{0,1\}^m$ where $m = k + 2\log\epsilon - 1/2$. Let $t = \lceil \log|\mathcal{H}| \rceil$. Then the above Ext is a (k,ϵ)-extractor for ND.

Proof: Without loss of generality, we assume that X is a flat k-source. Let $\epsilon = 2^{(1+m-k)/2}$. By Lemma 3, we have

$$\mathrm{Col}[(\boldsymbol{h},\boldsymbol{h}(X))] \le (1/(2^{t+m})(1+\epsilon^2).$$

Therefore,

$$\|(\boldsymbol{h},\boldsymbol{h}(X)) - U_{t+m}\|^2 = \mathrm{Col}[(\boldsymbol{h},\boldsymbol{h}(X))] - \frac{1}{2^{t+m}} \le \frac{\epsilon^2}{2^{t+m}}.$$

By Lemma 1, for any distribution P over $\{0,1\}^n$, we have

$$\begin{aligned}(ND(P,U_n))^2 &= \frac{\log e}{2}\cdot\sum_{j=1}^{\infty}\frac{1}{2j(2j-1)}\\ &\quad\times\left(\sum_{x\in\{0,1\}^n}\frac{|P(x)-2^{-n}|^{2j}}{(P(x)+2^{-n})^{2j-1}}\right)\\ &\le \frac{1}{2}\left(\sum_{x\in\{0,1\}^n}\frac{|P(x)-2^{-n}|^2}{(P(x)+2^{-n})}\right)\\ &\le 2^{n-1}\cdot\left(\sum_{x\in\{0,1\}^n}|P(x)-2^{-n}|^2\right)\\ &= 2^{n-1}\cdot\|P-U_n\|^2.\end{aligned}$$

Hence, we have

$$\begin{aligned}(ND((\boldsymbol{h},\boldsymbol{h}(X)),U_{(t+m)}) &\le \sqrt{2^{t+m-1}}\cdot\|(\boldsymbol{h},\boldsymbol{h}(X)) - U_{t+m}\|\\ &\le \frac{\epsilon}{\sqrt{2}}\\ &= \frac{1}{2^{(k-m)/2}}.\end{aligned}$$

This concludes that Ext is an extractor for ND. □

D. Expander Graphs

Similar to the Leftover Hash Lemma for ND, the expander-based extractor has the same property. Let G be a d-regular graph and M_G be its adjacency matrix. G is a λ-expander if the second largest eigenvalue of M_G is not greater than λ [1], [7]. We view a distribution as a vector. A random walk on λ-expander converges to the uniform distribution. Precisely, for any distribution P_n

$$\|{M_G}^k P_n - U_n\| \le \lambda^k\|P_n - U_n\|.$$

From the prior discussion, we get, for any distribution P_n on $\{0,1\}^n$

$$\begin{aligned}2^{1-n}(ND(M_GP_n,U_n))^2 &\le \|M_GP_n - U_n\|^2\\ &\le \lambda^2\left(\mathrm{Col}(P_n) - 2^{-n}\right).\end{aligned}$$

We define

$$\mathrm{Ext}_G : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^n$$

by setting $\mathrm{Ext}_G(x,y)$ to be the yth neighbor of x. Suppose X_n is a flat k-source and $-2\log\lambda \ge n - k - 2\log\epsilon$. Then we have

$$\begin{aligned}(ND(M_GX_n,U_n))^2 &\le 2^{n-1}\|M_GX_n - U_n\|^2\\ &\le 2^{n-1}\cdot\lambda^2\left(\mathrm{Col}(X_n) - 2^{-n}\right)\\ &\le \frac{\epsilon^2}{2}.\end{aligned}$$

Hence, we achieve the following expander-based extractor for ND.

Theorem 5: If G is a 2^t-regular λ-expander graph with $-2\log\lambda \ge n-k-2\log\epsilon$, then $\mathrm{Ext}_G : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^n$ is a (k,ϵ)-extractor for ND.

E. An Example That Does Not Carry Over to ND

From the preceding two subsections, we know that ND has a good bound in terms of the ℓ_2 norm for some special setting. Nevertheless,

TABLE I
DISTRIBUTION OF T_2

A	$\Pr[T_2 = A]$
00	0.389932
01	0.303991
10	0.201038
11	0.10504
$ND(T_2, U_2)$	0.073862
$\sum_{v\in\{0,1\}^2\setminus\{00\}} ND(T\cdot v, U_1)$	0.0689

ND is not linear in general. In this subsection, we give an example to show that L_1-distance has a more linear property. For SD metric, the parity lemma is as follows..

Lemma 4: (Parity Lemma) [12] For any t-bit random variable T

$$SD(T, U_t) \leq \sum_{v\in\{0,1\}^t\setminus\{0^t\}} SD(T\cdot v, U_1).$$

However, this statement is not true in general for ND. We find a counterexample. Let T_2 be the distribution as shown in Table I. By a simple calculation, we see that

$$ND(T_2, U_2) > \sum_{v\in\{0,1\}^2\setminus\{00\}} ND(T_2\cdot v, U_1).$$

Hence, the new metric ND does not hold for the parity lemma.

In order to find a general counterexample for $t \geq 2$, we define a distribution J_t on $\{0,1\}^t$ as

$$J_t = T_2 \circ \underbrace{U_1 \circ \cdots \circ U_1}_{t-2} = T_2 \circ U_{t-2}.$$

It is easy to get $ND(J_t, U_t) = ND(T_2, U_2)$. Next we want to show the following proposition.

Proposition 2:

$$\sum_{v\in\{0,1\}^t\setminus\{0^t\}} ND(J_t\cdot v, U_1) = \sum_{v\in\{0,1\}^2\setminus\{00\}} ND(T_2\cdot v, U_1).$$

Proof: Note that for any $t_2 \in \{0,1\}^2$ and for any nonzero vector $w \in \{0,1\}^{t-2}$

$$(t_2 \circ w)\cdot J_t \sim U_1.$$

Hence,

$$ND((t_2 \circ w)\cdot J_t, U_1) = 0.$$

Therefore,

$$\begin{aligned}&\sum_{v\in\{0,1\}^t\setminus\{0^t\}} ND(J_t\cdot v, U_1)\\ &\quad= \sum_{t_2\in\{0,1\}^2\setminus\{00\}} ND((T_2\circ U_{t-2})\cdot(t_2\circ 0^{t-2}), U_1)\\ &\quad= \sum_{t_2\in\{0,1\}^2\setminus\{00\}} ND(T_2\cdot t_2, U_1). \qquad \square\end{aligned}$$

In general, we get for any $t \geq 2$

$$ND(J_t, U_t) > \sum_{v\in\{0,1\}^t\setminus\{0^t\}} ND(J_t\cdot v, U_1).$$

TABLE II
COMPARISON BETWEEN SD AND ND

	SD	ND
Next-bit predictor	Applicable	Applicable but Factor Loss
Existence of extractor	Applicable	Applicable but Factor Loss
Leftover hash lemma	Applicable	Applicable
Expander graph	Applicable	Applicable
Parity lemma	Applicable	Non-Applicable

However, it is still possible that the parity lemma may exist for ND in a different form. Finally, we summarize the results of Section III in Table II.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for many useful comments.

REFERENCES

[1] N. Alon and J. Spencer, *The Probabilistic Method*, 2nd ed. New York: Wiley, 2000.

[2] Z. Bar-Yossef, “Sampling lower bounds via information theory,” in *Proc. 35th Annu. ACM Symp. Theory of Computing*, San Diego, CA, 2003, pp. 335–344.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[4] D. M. Endres and J. E. Schindelin, “A new metric for probability distributions,” *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1858–1860, Jul. 2003.

[5] R. Impagliazzo and D. Zuckerman, “How to recycle random bits,” in *Proc. 30th IEEE Symp. Foundations of Computer Science*, NC, Oct. 1989, pp. 248–253.

[6] J. Lin, “Divergence measures based on the shannon entropy,” *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, Jan. 1991.

[7] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[8] N. Nisan and D. Zuckerman, “Randomness is linear in space,” *J. Comp. Syst. Sci.*, vol. 52, no. 1, pp. 43–52, Feb. 1996.

[9] D. R. Stinson, “Universal hash families and the leftover hash lemma, and applications to cryptography and computing,” *J. Combin. Math. Combin. Comput*, vol. 42, pp. 3–31, 2002.

[10] F. Topsøe, “Some inequalities for information divergence and related measures of discrimination,” *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1602–1609, Jul. 2000.

[11] L. Trevisan, “Construction of extractors using pseudorandom generators,” in *Proc. 31st ACM Symp. Theory of Computing*, Atlanta, GA, 1999, pp. 141–148.

[12] U. Vazirani, “Strong communication complexity of generating quasirandom sequences from two communicating semi-random sources,” *Combinatorica*, vol. 7, no. 4, pp. 375–392, 1987.

[13] A. C. Yao, “Theory and applications of trapdoor functions,” in *Proc. 23rd Annu. IEEE Symp. Foundations of Computer Science*, Chicago, IL, Nov. 1982, pp. 80–91.

[14] R. W. Yeung, *A First Course in Information Theory*. New York: Kluwer Academic/Plenum, 2002.