Int. Computer Symposium, Dec. 15-17, 2004, Taipei, Taiwan.

On the Jensen-Shannon Divergence and Variational Distance

Shi-Chun Tsai
Dept of Computer Science
and Info Engineering

Wen-Guey Tzeng
Dept of Computer
and Info Science

Hsin-Lung Wu
Dept of Computer Science
and Info Engineering

National Chiao-Tung University

sctsai@csie.nctu.edu.tw *tzeng@cis.nctu.edu.tw* *hsinlung@csie.nctu.edu.tw*

***Abstract**-We study the distance measures between two probability distributions via two different distance metrics, a new metric induced from Jensen-Shannon Divergence[4] and the well known L_1 metric. First we show that the bounds between these two distance metrics are tight for some particular distributions. Then we show that the L_1 distance of a binomial distribution does not imply the entropy power inequality for the binomial family, proposed in [5].*

*Moreover, 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[**?**] and the construction of expander graph based extractor. Finally we show that the useful parity lemma [12] in studying pseudo-randomness does not hold in the new metric.*

Keywords: Jensen-Shannon Divergence, variational distance, extractors.

1 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 pseudo-randomness 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[7]. For breaking the condition of absolute continuity of Kullback divergence. These researches are information-theoretic. We will use Jensen-Shannon Divergence to investigate some computational complexity issues.

In this paper, we bound variational distance SD by the new distance ND and show that the bound is tight. Then we show that it is unlikely to prove entropy power inequality for binomial family via the bound from L_1 metric. Moreover, 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], leftover hash lemma[**?**] and the construction of expander graph based extractors. Finally we show that the useful parity lemma [12] in studying pseudo-randomness does not hold in the new metric.

	SD	ND
Entropy power inequality for binomial family	Non-Applicable	Applicable
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

Table 1: Comparison between SD and ND

2 Preliminaries

We use $[n]$ to denote the set $\{1, 2, .., n\}$. The base of log function is 2. For any distribution X with sample space $\Omega_n = \{\omega_1, .., \omega_n\}$, define the entropy of X to be $H(X) = \sum_{i=1}^{n} - \Pr[X = \omega_i] \log \Pr[X = \omega_i]$. For every positive integer m, U_m 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.

Let Π be the set of distributions whose sample space is Ω_n. We use a metric function to measure the distance between two distributions. A metric function satisfies the following properties.

Definition 1 *We say that a function $F : \Pi \times \Pi \to [0, 1]$ is a metric if* (**a**)$F(P, Q) = 0$ *if and only if* $P = Q$, (**b**)$F(P, Q) = F(Q, P)$, *and* (**c**)*for any* $P, Q, R \in \Pi$, $F(P, Q) \leq F(P, R) + F(R, Q)$.

We could easily prove that variational distance is a metric. The following facts are useful in this paper.

Fact 1 *Function SD is a metric where* $SD(P, Q) = \frac{1}{2} \sum_{i=1}^{n} | \Pr[P = \omega_i] - \Pr[Q = \omega_i]|$.

Fact 2 $\ln 2 = \sum_{j=1}^{\infty} \frac{1}{2j(2j-1)}$.

3 A tight relation between ND and SD

Let P and Q be two distributions with the same probability space and T be a 0-1 random variable with $\Pr[T = 0] = 1/2$ and independent of P and Q. Define the following distribution:

$$Z_{PQ} = \begin{cases} P & \text{if } T = 0 \\ Q & \text{if } T = 1. \end{cases}$$

Definition 2 *The Jensen-Shannon Divergence is* $(H(Z_{PQ}) - (H(P) + H(Q))/2)$. *$ND$ is defined as*

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

Endres and Schindelin proved that ND is a metric[4]. Suppose $P =< p_1, \cdots, p_n >$ and $Q =< q_1, \cdots, q_n >$ where $p_i = \Pr[P = \omega_i]$ and $q_i = \Pr[Q = \omega_i]$ for $1 \leq i \leq n$. We need a lemma proved by Topsøe [10].

Lemma 1 *[10] For any distributions P and Q in Π,*

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

We reprove the following in a more direct way.

Theorem 1 *[10]* $\sqrt{SD(P, Q)} \geq ND(P, Q) \geq \sqrt{\frac{(1+SD(P,Q)) \log(1+SD(P,Q)) + (1-SD(P,Q)) \log(1-SD(P,Q))}{2}}$.

Actually, the above bounds are tight. For the left-hand-side inequality, we consider the following two distributions: $P =< \epsilon, \underbrace{\frac{1-\epsilon}{n-2}, \cdots, \frac{1-\epsilon}{n-2}}_{n-2}, 0 >$ and $Q =< 0, \underbrace{\frac{1-\epsilon}{n-2}, \cdots, \frac{1-\epsilon}{n-2}}_{n-2}, \epsilon >$. 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: $P =< \underbrace{\frac{1+\epsilon}{2n}, \cdots, \frac{1+\epsilon}{2n}}_{n}, \underbrace{\frac{1-\epsilon}{2n}, \cdots, \frac{1-\epsilon}{2n}}_{n} >$ and $Q =< \underbrace{\frac{1-\epsilon}{2n}, \cdots, \frac{1-\epsilon}{2n}}_{n}, \underbrace{\frac{1+\epsilon}{2n}, \cdots, \frac{1+\epsilon}{2n}}_{n} >$. Clearly $SD(P,Q) = \epsilon$. And we have:

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

Therefore the right-hand side is a tight bound.

4 Advantage of ND

In this section we show that Theorem 1 does not help to prove the entropy power inequality for the binomial family in [5]. This shows that ND is more suitable than SD in this case. The following facts will be handy in the rest of this section.

Fact 3 *[3] Suppose P and Q are two distributions on $\mathcal{A}$. Let $\mathcal{B} = \{x \in \mathcal{A} : P(x) \geq Q(x)\}$. Then $SD(P,Q) = \Pr[P \in \mathcal{B}] - \Pr[Q \in \mathcal{B}]$.*

Fact 4 *[2] $\binom{n}{\lceil \frac{n}{2} \rceil} < 2^n \sqrt{\frac{2}{\pi}} \sqrt{\frac{2n+1}{2n^2}}$.*

Let $X_1, \cdots, X_n, \cdots$ be an i.i.d. random process where each $X_i \sim U_1$. Let $Y_n = \sum_{i=1}^{n} X_i$. Then Y_n is a binomial distribution with parameters n and $\frac{1}{2}$. The entropy power inequality for the binomial family states that: for any $m, n \geq 1$, $2^{2H(Y_n)} + 2^{2H(Y_m)} \leq 2^{2H(Y_n+Y_m)}$. An easy observation is that if $\frac{2^{2H(Y_n)}}{n}$ is increasing in n then the power inequality holds. Hence we just need to show that $\frac{2^{2H(Y_n)}}{n}$ is increasing. It is sufficient to prove the following lower bound: $H(Y_{n+1}) - H(Y_n) \geq \frac{1}{2}\log\frac{n+1}{n}$. Denote P_Y as the probability distribution of Y. It is clear that $P_{Y_{n+1}} = \frac{P_{Y_n} + P_{Y_n+1}}{2}$. By the definition of Jensen-Shannon Divergence, we have $H(Y_{n+1}) = H(Y_n+1)/2 + H(Y_n)/2 + ND^2(P_{Y_n}, P_{Y_n+1})$. Note that $H(Y_n) = H(Y_n+1)$. Hence we have $H(Y_{n+1}) = H(Y_n) + ND^2(P_{Y_n}, P_{Y_n+1})$. The following has been proved by Harremoës and Vignat[5]

$$ND^2(P_{Y_n}, P_{Y_n+1}) \geq \frac{1}{2}\log\frac{n+1}{n}. \tag{1}$$

Thus we obtain a lower bound for $H(Y_{n+1}) - H(Y_n)$ via ND.

We may hope that Theorem 1 will help us to prove Inequality (1). However we cannot prove it via Theorem 1. In fact we can prove the following inequality for large n

$$(2\ln 2)(SD(P_{Y_n}, P_{Y_n+1}))^2 < \frac{1}{n} - \frac{1}{2n^2}. \tag{2}$$

This implies (as in the proof of Theorem 1) that

$$\sum_{j=1}^{\infty} \frac{1}{j(2j-1)}(SD(P_{Y_n}, P_{Y_n+1}))^{2j} < \ln\frac{1+n}{n} \tag{3}$$

Inequality (3) tells us that Theorem 1 does not help us prove Inequality (1). Finally we show that Inequality (2) is correct for large n. We can view P_{Y_n} and P_{Y_n+1} as two distributions on $\{0, 1, \cdots, n+1\}$.

By Fact 3 and 4 we have $SD(P_{Y_n}, P_{Y_n+1}) = 2^{-n}\binom{n}{\lceil \frac{n}{2} \rceil} < \sqrt{\frac{2}{\pi}}\sqrt{\frac{2n+1}{2n^2}}$. It is easy to check that the following inequalities: $(2\ln 2)(SD(P_{Y_n}, P_{Y_n+1}))^2 < \frac{1}{n} - \frac{1}{2n^2}$.

5 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 [8]. In this section we illustrate that several important results in randomized computation based on SD carry over to ND. While we also show a non-applicable case.

5.1 Distinguisher v.s. 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 3 *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*

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

Definition 4 *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 $|Pr_{x \leftarrow D_n}[G(x_1, \cdots, x_{i-1}) = x_i]| \geq \epsilon$.*

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 $\frac{\epsilon}{n}$-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 $Pr[G(x_1, \cdots, x_{i-1}) = x_i] \geq \frac{\epsilon^2}{n}$, where $x_1, \cdots, x_i$ 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 $\frac{\epsilon^2}{n}$-good next-bit predictor G for D_n. □

5.2 Extractors

We continue to show the existence of extractors under the setting of ND with some appropriate parameters. Similar to the definition of extractor [9], we have the following definition.

Definition 5 *$EXT : \{0,1\}^n \times \{0,1\}^t \rightarrow \{0,1\}^m$ is called a (k, ϵ)-extractor for ND if for every k-source D_n, $ND(EXT(D_n, U_t), U_m) \leq \epsilon$.*

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 $EXT : \{0,1\}^n \times \{0,1\}^t \rightarrow \{0,1\}^m$ for ND with $t = \log n - k - 4\log\epsilon + O(1)$ and $m = k + t + 4log\epsilon - O(1)$.*

Proof. We prove the proposition by the probabilistic method [1, 8]. 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 $T : \{0,1\}^m \rightarrow \{0,1\}$ 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 $Supp(Q) = \{x : Q(x) > 0\}$. For each $x \in Supp(Q)$, the distribution of $f(x)$ is the same as U_m. Also $\{f(x) : x \in Supp(Q)\}$ is a set of random variables which are i.i.d. For each boolean function $T : \{0,1\}^m \rightarrow \{0,1\}$, $\{T(f(x)) : x \in Supp(Q)\}$ is also a set of 0-1 random variables which are i.i.d. and $Exp[T(f(x))] = \frac{|\{z:T(z)=1\}|}{2^m} = \Pr[(U_m)_T = 1]$. By the Chernoff Bound [1, 8], $\Pr[|\frac{\sum_{x \in Supp(Q)} T(f(x))}{2^{k+t}} - \frac{|\{z:T(z)=1\}|}{2^m}| > \epsilon^2] < 2^{-\Omega(2^{k+t}\epsilon^4)}$. By Theorem 1, we can get $\Pr[ND(f(Q), U_m) > \epsilon] \leq \Pr[SD(f(Q), U_m) > \epsilon^2] \leq \Pr[\exists T, SD(f(Q)_T, (U_m)_T) > \epsilon^2] < 2^{2^m} \cdot 2^{-\Omega(2^{k+t} \cdot \epsilon^4)}$. □

The probability that f is not a good extractor for some flat k-source is at most $\binom{2^n}{2^k} \cdot 2^{2^m} \cdot 2^{-\Omega(2^{k+t} \cdot \epsilon^4)} < 1$. 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.

5.3 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 ℓ_2 norm. This bound can help us prove some results about extractors for ND.

Definition 6 *[6] $\mathcal{H} = \{h : \mathcal{D} \to \mathcal{R}\}$ 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 $\Pr_{h \leftarrow \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{|\mathcal{R}|} + \frac{1}{|\mathcal{D}|}$.*

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 *[6] Suppose $\mathcal{H}$ is almost universal, X is a flat k-source on $\{0,1\}^n$, and $\mathbf{h}$ is a random function drawn from $\mathcal{H}$. Then $SD\left((\mathbf{h}, \mathbf{h}(X)), U_{t+m}\right) \leq 2^{(m-k)/2}$.*

Define $Col[(\mathbf{h}, \mathbf{h}(X))] = \Pr[(\mathbf{h}, \mathbf{h}(X)) = (\mathbf{h}', \mathbf{h}'(X'))]$ where $\mathbf{h}', X'$ are i.i.d. to $\mathbf{h}, X$, respectively. The crucial part of the proof of Theorem 3 is to show the following lemma.

Lemma 3 *[6]*

$$Col[(\mathbf{h}, \mathbf{h}(X))] \leq (1 + 2^{(1+m-k)})/(2^{t+m}).$$

Define $Ext : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^{t+m}$ by $Ext(x, h) = (h, h(x))$. 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 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 $Col[(\mathbf{h}, \mathbf{h}(X))] \leq \frac{1}{2^{t+m}}(1+\epsilon^2)$. Therefore $\|(\mathbf{h}, \mathbf{h}(X)) - U_{t+m}\|^2 = Col[(\mathbf{h}, \mathbf{h}(X))] - \frac{1}{2^{t+m}} \leq \frac{\epsilon^2}{2^{t+m}}$. By the proof of Theorem 1, for any distribution P over $\{0,1\}^n$, we have $(ND(P, U_n))^2 \leq \frac{1}{2}\left(\sum_{x \in \{0,1\}^n} \frac{|P(x) - 2^{-n}|^2}{(P(x) + 2^{-n})}\right) = 2^{n-1} \cdot \|P - U_n\|^2$. Hence we have $(ND((\mathbf{h}, \mathbf{h}(X)), U_{(t+m)}) \leq \frac{1}{2^{(k-m)/2}}$. This concludes that Ext is an extractor for ND. □

5.4 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, 8]. We view a distribution as a vector. A random walks on λ-expander converges to the uniform distribution. Precisely, for any distribution P_n, $\|{M_G}^k P_n - U_n\| \leq \lambda^k \|P_n - U_n\|$. From the prior discussion, we get, for any distribution P_n on $\{0,1\}^n$, $2^{1-n}(ND(M_G P_n, U_n))^2 \leq \|M_G P_n - U_n\|^2 \leq \lambda^2 \left(Col(P_n) - 2^{-n}\right)$. We define $Ext_G : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^n$ by setting $Ext_G(x, y)$ to be the y-th neighbor of x. Suppose X_n is a flat k-source and $-2\log\lambda \geq n - k - 2\log\epsilon$. Then we have $(ND(M_G X_n, U_n))^2 \leq 2^{n-1}\|M_G X_n - U_n\|^2 \leq 2^{n-1} \cdot \lambda^2 \left(Col(X_n) - 2^{-n}\right) \leq \frac{\epsilon^2}{2}$ Hence we achieve the following expander-based extractor for ND.

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

5.5 An example that doesn't carry over to ND

In the previous 2 subsections, we know that ND has a good bound in terms of ℓ_2 norm for some special setting. Nevertheless ND is not linear in general. In this subsection, we give an example to show that L_1-distance has more linear property. For SD metric, the parity lemma is as following.

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

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

Table 2: Distribution of T_2

as shown in Table 2. 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 $J_t = T_2 \circ U_{t-2}$. 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 $w \in \{0,1\}^{t-2}$, $(t_2 \circ w) \cdot J_t = U_1$. Hence $ND((t_2 \circ w) \cdot J_t, U_1) = 0$. Therefore $\sum_{v \in \{0,1\}^t \setminus \{0^t\}} ND(J_t \cdot v, U_1) = \sum_{t_2 \in \{0,1\}^2 \setminus \{00\}} ND\left((T_2 \circ U_{t-2}) \cdot (t_2 \circ 0^{t-2}), U_1\right) = \sum_{t_2 \in \{0,1\}^2 \setminus \{00\}} ND(T_2 \cdot t_2, U_1)$. □

In general we get, for any $t \geq 2$, $ND(J_t, U_t) > \sum_{v \in \{0,1\}^t \setminus \{0^t\}} ND(J_t \cdot v, U_1)$. However, it is still possible that the parity lemma may exist for ND in a different form.

References

[1] N. Alon and J. Spencer. The Probabilistic Method, 2nd Ed, John Wiley & Sons, Inc., 2000.

[2] Kenneth. P. Bogart. Introductory Combinatorics. Third Edition. Academic Press. 2000.

[3] T. M. Cover and J. A. Thomas. Elements of Information Theory. John Wiley & Sons, Inc., 1991.

[4] Dominik M. Endres and Johannes E. Schindelin. A New Metric for Probability Distributions. IEEE Transaction on Information Theory, vol 49, pp.1858-60. July 2003.

[5] Peter Harremoës and Christophe Vignat. An Entropy Power Inequality for th Binomial Family. Journal of Inequalities in Pure and Applied Mathematics. Vol 4, Issue 5, Article 93, 2003.

[6] Impagliazzo, R. and D. Zuckerman, How to Recycle Random Bits, Proceedings of 30th IEEE Symposium on the Foundations of Computer Science, Research Triangle Park, NC, October 1989, pp. 248-253.

[7] Jianhua Lin. Divergence Measures Based on the Shannon Entropy. IEEE Transaction on Information Theory, vol 37, No. 1. pp.145-151. January 1991.

[8] R. Motwani and P. Raghavan. Randomized Algorithms, Cambridge University Press, 1995.

[9] Noam Nisan and David Zuckerman. Randomness is linear in space. Journal of Computer and System Sciences, 52(1):43-52, February 1996.

[10] F. Topsøe. Some inequalities for information divergence and related measures of discrimination. IEEE Transaction on Information Theory, vol IT-46 no.4, pp.1602-1609. July 2000.

[11] L. Trevisan. Construction of extractors using pseudorandom generators. In Proceedings of the 31st ACM Symposium on Theory of Computing, 1999.

[12] U. Vazirani. Strong Communication Complexity of Generating Quasi-Random Sequences from Two Communicating Semi-Random Sources. Combinatorica, 7(4):375-392, 1987.

[13] Andrew C. Yao. Theory and applications of trapdoor functions. In 23rd Annual Symposium on Foundations of Computer Science, pages 80-91, Chicago, Illinois, 3-5 November 1982. IEEE.