12. Hellinger distance

Lecturer: Prahladh Harsha *Scribe: Girish Varma*

In this lecture, we will introduce a new notion of distance between probability distributions called *Hellinger distance*. Using some of the nice properties of this distance, we will generalize the fooling set argument for deterministic protocols to the randomized setting. We will then use this to prove a $\Omega(n)$ lower bound for the communication complexity of Disjointness. We will also see how this proof easily extends to the multi party setting, thereby proving Theorem 5.7 from Lecture 5.

12.1 Hellinger Distance

Let $P = \{p_i\}_{i\in[n]}, Q = \{q_i\}_{i\in[n]}$ be two probability distributions supported on $[n]$. A natural way of defining a distance between them is to consider the ℓ_1-distance between the probability vectors P and Q.

$$\|P - Q\|_1 = \sum_{i\in[n]} |p_i - q_i|.$$

The *total variation distance*, denoted by $\Delta(P, Q)$ (and sometimes by $\|P - Q\|_{TV}$), is half the above quantity. It is an easy exercise to check that

$$\Delta(P, Q) = \max_{S\subseteq[n]} |P(S) - Q(S)|. \tag{12.1.1}$$

Because of the above equality, this is also referred to as the *statistical distance*.

Taking the ℓ_1 norm of the difference made sense because P and Q where unit vectors according to the ℓ_1 norm. Since $\sqrt{P} = (\sqrt{p_1}, \sqrt{p_2}, \ldots \sqrt{p_n})$ is a unit vector according to ℓ_2 norm, we can also consider the ℓ_2 norm of the difference of the square root vectors.

Definition 12.1 (Hellinger Distance)**.** *For probability distributions $P = \{p_i\}_{i\in[n]}, Q = \{q_i\}_{i\in[n]}$ supported on $[n]$, the* Hellinger distance *between them is defined as*

$$h(P, Q) = \frac{1}{\sqrt{2}} \cdot \|\sqrt{P} - \sqrt{Q}\|_2.$$

By definition, the Hellinger distance is a metric satisfying triangle inequality. The $\sqrt{2}$ in the definition is for ensuring that $h(P, Q) \leq 1$ for all probability distributions. It is closely related to a quantity known as Fidelity or the Bhattacharya coefficient of two probability distributions $F(P, Q) = \sum_{i\in[n]} \sqrt{p_i q_i}$ by the relation:

$$h^2(P, Q) = 1 - F(P, Q).$$

12.1.1 Properties of Hellinger distance

Lemma 12.2 (Hellinger vs. total variation)**.**

$$h^2(P,Q) \le \Delta(P,Q) \le \sqrt{h^2(P,Q)(2-h^2(P,Q)} \le \sqrt{2}h(P,Q)$$

Proof. For the first inequality,

$$\begin{aligned} h^2(P,Q) &= \frac{1}{2}\sum_i |\sqrt{p_i}-\sqrt{q_i}||\sqrt{p_i}-\sqrt{q_i}| \le \frac{1}{2}\sum_i |\sqrt{p_i}-\sqrt{q_i}|(\sqrt{p_i}+\sqrt{q_i}) \\ &\le \frac{1}{2}\sum_i |p_i - q_i| = \Delta(P,Q). \end{aligned}$$

For the last two inequalities,

$$\begin{aligned} \Delta^2(P,Q) &= \frac{1}{4}\left(\sum_{i\in[n]} |p_i-q_i|\right)^2 = \frac{1}{4}\left(\sum_{i\in[n]} \left(\sqrt{p_i}-\sqrt{q_i}\right)\left(\sqrt{p_i}+\sqrt{q_i}\right)\right)^2 \\ &\le \frac{1}{4}\left(\sum_{i\in[n]} \left(\sqrt{p_i}-\sqrt{q_i}\right)^2\right)\left(\sum_{i\in[n]} \left(\sqrt{p_i}+\sqrt{q_i}\right)^2\right) \qquad \text{[By Cauchy Schwarz]} \\ &\le \frac{1}{2}\cdot h^2\left(P,Q\right)\cdot\left(2+2\sum_{i\in[n]}\sqrt{p_i}\sqrt{q_i}\right) \\ &\le h^2\left(P,Q\right)\cdot\left(2-h^2\left(P,Q\right)\right) \le \sqrt{2}h\left(P,Q\right). \end{aligned}$$

□

Cut and paste property: In the fooling set argument, we saw that if inputs (x,y) and (x',y') have the same transcript in a deterministic communication protocol, then (x',y) and (x,y') must have the same transcript. This rectangle property can be extended to private coins randomized protocols using Hellinger distance in the follows sense: if the transcript distributions for inputs (x,y) and (x',y') are close in Hellinger distance, then so are the transcript distributions for (x',y) and (x,y').

Lemma 12.3 (Cut-and-Paste)**.** *Let $\mathcal{P}$ be a randomized private coins protocol and $\Pi_{x,y}$ denote the (randomized) transcript on input x,y. Then,*

$$h^2(\Pi_{x,y},\Pi_{x',y'}) = h^2(\Pi_{x',y},\Pi_{x,y'}).$$

Proof. We can think of a randomized private coin protocol working on input (x,y) as a deterministic protocol on the extended inputs $((x,R_A),(y,R_B))$, where the additional inputs R_A and R_B are chosen according to the suitable private coins distribution. From the rectangle property of deterministic protocols, we have that for any fixed transcript τ, the set of extended inputs that gives rise to it form a rectangle, say $\mathsf{Rect}_\tau = S_\tau \times T_\tau$. Now,

let's consider the probability that transcript τ arises for inputs x and y.

$$\begin{aligned}\Pr_{R_A,R_B}[\Pi(x,y,R_A,R_B)=\tau] &= \Pr_{R_A,R_B}[((x,R_A),(y,R_B)) \in \mathsf{Rect}_\tau] \\ &= \Pr_{R_A,R_B}[(x,R_A)\in S_\tau \text{ and } (y,R_B)\in T_\tau] \\ &= \Pr_{R_A}[(x,R_A)\in S_\tau]\cdot \Pr_{R_B}[(y,R_B)\in T_\tau].\end{aligned}$$

This splitting of probabilities follows from the independence of Alice and Bob's private coins R_A and R_B and is used to proved the lemma as follows.

$$\begin{aligned}&1-h^2(\Pi_{x,y},\Pi_{x',y'}) = F(\Pi_{x,y},\Pi_{x',y'}) \\ &= \sum_\tau \sqrt{\Pr_{R_A,R_B}[\Pi_{x,y}=\tau]\cdot \Pr_{R_A,R_B}[\Pi_{x',y'}=\tau]} \\ &= \sum_\tau \sqrt{\Pr_{R_A}[(x,R_A)\in S_\tau]\cdot \Pr_{R_B}[(y,R_B)\in T_\tau]\cdot \Pr_{R_A}[(x',R_A)\in S_\tau]\cdot \Pr_{R_B}[(y',R_B)\in T_\tau]} \\ &= \sum_\tau \sqrt{\Pr_{R_A}[(x,R_A)\in S_\tau]\cdot \Pr_{R_B}[(y',R_B)\in T_\tau]\cdot \Pr_{R_A}[(x',R_A)\in S_\tau]\cdot \Pr_{R_B}[(y,R_B)\in T_\tau]} \\ &= \sum_\tau \sqrt{\Pr_{R_A,R_B}[\Pi_{x,y'}=\tau]\cdot \Pr_{R_A,R_B}[\Pi_{x',y}=\tau]} \\ &= F(\Pi_{x,y'},\Pi_{x',y}) = 1-h^2(\Pi_{x,y'},\Pi_{x',y}).\end{aligned}$$

□

The above cut-and-paste lemma can be extended to communication protocols for t parties.

Lemma 12.4 (multiparty cut-and-paste)**.** *For any $v \in \{x_1,y_1\}\times\{x_2,y_2\}\cdots\times\{x_t,y_t\}$*

$$h^2(\Pi_{x_1,x_2,\ldots x_t},\Pi_{y_1,y_2,\ldots y_t}) = h^2(\Pi_v,\Pi_{\overline{v}}).$$

Lemma 12.5 (Hellinger vs. Information [Lin91])**.** *Let Z be a random variable taking values in $\{z_1,z_2\}$ equally likely and Π a randomized function of Z. Then,*

$$I[Z:\Pi(Z)] \geq h^2(\Pi_{z_1},\Pi_{z_2}).$$

A proof of a slightly weaker theorem is presented in Appendix A.

12.2 Lower bound for Disjointness

In this section, we will prove the $\Omega(n)$ lower bound for the randomized private coins communication complexity of Disjointness, using the above properties of Hellinger distance. Recall that

$$\mathsf{DISJ}(x,y) = \bigwedge_i \overline{x_i}\vee\overline{y_i} = \bigwedge_i \mathsf{NAND}(x_i,y_i).$$

Let's quickly recall the steps in the proof of the disjointness lower bound from last lecture.

Suppose there exists a $(1/2-\varepsilon)$-error private coin randomized protocol Π for computing the Disjointness problem of length at most δn for some $\delta > 0$. Define two distribution η_A and η_B on $\{0,1\}^2$ (in fact, on the YES instances of NAND) as follows.

$$\Pr_{(X,Y)\sim\eta_A}[(X,Y)=(1,0)] = \Pr_{(X,Y)\sim\eta_A}[(X,Y)=(0,0)] = \frac{1}{2}$$
$$\Pr_{(X,Y)\sim\eta_B}[(X,Y)=(0,1)] = \Pr_{(X,Y)\sim\eta_B}[(X,Y)=(0,0)] = \frac{1}{2}$$

For every $\sigma \in \{A,B\}^n$, we define joint random variables (X^σ, Y^σ) with distribution μ_σ on $(\{0,1\}^n)^2$ as follows: for each $i \in [n]$ independently do the following, if $\sigma_i = A$, set $(X_i^\sigma, Y_i^\sigma) \sim \eta_A$ and otherwise (i.e., $\sigma_i = B$), set $(X_i^\sigma, Y_i^\sigma) \sim \eta_B$. Using, sub-additivity of mutual information and independence of (X_i^σ, Y_i^σ) across the different i's, we showed that

$$\delta n \geq |\Pi(X^\sigma, Y^\sigma)| \geq I[X^\sigma, Y^\sigma : \Pi(X^\sigma, Y^\sigma)] \geq \sum_i I[X_i^\sigma, Y_i^\sigma : \Pi(X^\sigma, Y^\sigma)].$$

Averaging over i's and all possible σ's we get,

$$\delta \geq \mathbb{E}_\sigma E_k I[X_k^\sigma Y_k^\sigma : \Pi(X^\sigma, Y^\sigma)] = E_k E_{\sigma_{-k}} E_{\sigma_k} I[X_k^\sigma Y_k^\sigma : \Pi].$$

Hence, there exists a k and a σ_{-k}, such that $\mathbb{E}_{\sigma_k} I[X_k^\sigma Y_k^\sigma : \Pi] \leq \delta$. Expanding this expectation, we obtain

$$\frac{1}{2}\left(I_A[X_k Y_k : \Pi] + I_B[X_k Y_k : \Pi]\right) \leq \delta, \tag{12.2.1}$$

where $I_A[\cdot,\cdot]$ denotes the mutual information when when the k-coordinates are chosen according to η_A and the remaining coordinates are chosen as dictated by the σ_{-k} that we had fixed earlier in the proof (similarly for $I_B[\cdot,\cdot]$).

This gives a protocol π for computing the NAND function which works as follows: On input x and y, Alice and Bob construct n bit inputs X and Y for DISJ_n function from x and y respectively as follows: Alice and Bob set the k-bit of X and Y to be x and y respectively (i.e., $X_k = x$ and $Y_k = y$). For each $i \neq k$, $\sigma_{-k}|_i$ tells if Alice or Bob is active. If Alice is active (ie., $\sigma_{-k}(i) = A$), then Alice sets X_i with equal probability to 0 or 1, while Bob sets Y_i to be 0. Similarly, if Bob is active, then Bob sets Y_i with equal probability to 0 or 1, while Alice sets X_i to be 0. Observe, that all of this can be done by Alice and Bob independently using their private randomness and the knowledge of k and σ_{-k}. They, then run the protocol Π on this input (X,Y). Since $\mathsf{DISJ}_n(X,Y) = \mathsf{NAND}(x,y)$ and Π is a protocol that computes DISJ_n correctly on every input with error at most $1/2-\varepsilon$, we have that π is a protocol that computes NAND correctly on every input with error at most $1/2-\varepsilon$. Rewriting (12.2.1) in terms of protocol π, we have

$$I[Z : \pi_{Z,0}] + I[Z : \pi_{0,Z}] \leq 2\delta, \tag{12.2.2}$$

where Z is a random bit that takes 0 and 1 with equal probability.

We can now complete the lower bound using the properties of the Hellinger distance proved in the beginning of the lecture.

$$
\begin{aligned}
2\delta &\geq I[Z : \pi_{Z,0}] + I[Z : \pi_{0,Z}] && \text{[from (12.2.2)]} \\
&\geq h^2(\pi_{0,0}, \pi_{1,0}) + h^2(\pi_{0,0}, \pi_{0,1}) && \text{[from Lemma 12.5]} \\
&\geq \frac{1}{2} \cdot (h(\pi_{0,0}, \pi_{1,0}) + h(\pi_{0,0}, \pi_{0,1}))^2 && \text{[By Cauchy-Schwarz]} \\
&\geq \frac{1}{2} \cdot h^2(\pi_{1,0}, \pi_{0,1}) && \text{[triangle inequality, since } h \text{ is a metric]} \\
&= \frac{1}{2} \cdot h^2(\pi_{0,0}, \pi_{1,1}) && \text{[by Cut-and-Paste lemma 12.3]} \\
&\geq \frac{1}{4} \cdot \Delta^2(\pi_{0,0}, \pi_{1,1}) && \text{[by Lemma 12.2]} \\
&\geq \varepsilon^2
\end{aligned}
$$

The last inequality follows from the fact that NAND takes different answers on inputs $(1,1)$ and $(0,0)$. More precisely, if the protocol was correct with probability $1/2 + \varepsilon$, using Equation (12.1.1) we obtain:

$$
\begin{aligned}
\Delta(\pi_{1,1}, \pi_{0,0}) &\geq |\Pr[\text{Output of transcript } \pi_{1,1} = 0] - \Pr[\text{Output of transcript } \pi_{0,0} = 0]| \\
&\geq \left| \left(\frac{1}{2} + \varepsilon\right) - \left(\frac{1}{2} - \varepsilon\right) \right| \\
&\geq 2\varepsilon
\end{aligned}
$$

Hence, $\delta \geq \varepsilon^2/2$. We have thus, proved the following theorem.

Theorem 12.6 (Disjointness lower bound [KS92, Raz92, BJKS04])**.**

$$R_{1/2-\varepsilon}(\mathsf{DISJ}_n) \geq \varepsilon^2 n/2.$$

12.3 Lower bound for Multi-party Disjointness

In this section, we will generalize the proof for disjointness to the multi-party setting. Recall the promise problem of $\mathsf{UDISJ}_{n,t}$ from Lecture 5, given by:

$$
\begin{aligned}
\mathsf{YES} &= \{(x_1, x_2, \ldots x_t) \in \{0,1\}^{nt} \mid \forall i \neq j, x_i, x_j \text{ are pairwise disjoint } \} \\
\mathsf{NO} &= \{(x_1, x_2, \ldots x_t) \in \{0,1\}^{nt} \mid \exists a \in [n],\ \forall i \neq j,\ x_i \cap x_j = \{a\}\}
\end{aligned}
$$

Note that we are using x_i to denote subsets of $[n]$ as well as the characteristic vectors of these subsets.

As in the last section, we will start with the observation that

$$\mathsf{UDISJ}_{n,t}(x_1, \ldots, x_t) = \bigwedge_{i=1}^{n} \left(\bigvee_{j=1}^{t} \overline{x_{j,i}} \right) = \bigwedge_{i=1}^{n} \mathsf{NAND}_t(x_{1,i}, x_{2,u}, \ldots, x_{t,i}).$$

Just as UDISJ is a promise problem, the t-wise NAND_t is also a promise problem whose only NO instance is the all 1's vector $\mathbf{1}$ and YES instances are the unit vectors $\mathbf{e_i}$ and the all zeros vector $\mathbf{0}$. Here, $\mathbf{e_i}$ represents the unit vector with 1 in the i^{th} coordinate.

We consider the number-in-hand and broadcast model and prove the following theorem.

Theorem 12.7 ([BJKS04]). *The communication complexity of* $\mathsf{UDISJ}_{n,t}$ *under number in hand and broadcast communication is* $\Omega(n/t^2)$.

As in the case of DISJ, we will show that communication complexity of $\mathsf{UDISJ}_{n,t}$ is at least n times the information complexity of multi-party NAND_t, which we will lower bound by $\Omega(1/t^2)$, again using properties of Hellinger distance.

Proof. Suppose there exists a $(1/2-\varepsilon)$-error private coin multi-party NIH randomized broadcast protocol Π for computing $\mathsf{UDISJ}_{n,t}$ of total broadcast length at most δn for some $\delta > 0$. Define t distributions $\eta_i, i \in [t]$ on $\{0,1\}^t$ (in fact, on the YES instances of NAND_t) as follows.

$$\Pr_{(X_1,\ldots,X_t)\sim\eta_i}[(X_1,X_2,\ldots,X_t)=\mathbf{e_i}] = \Pr_{(X_1,\ldots,X_t)\sim\eta_i}[(X_1,X_2,\ldots,X_t)=\mathbf{0}] = \frac{1}{2}.$$

For every $\sigma \in \{A_1,\ldots,A_t\}^n$, we define joint random variables $(X_1^\sigma,\ldots,X_t^\sigma)$ with distribution μ_σ on $(\{0,1\}^n)^t$ as follows: for each $i \in [n]$ independently do the following, if $\sigma_i = A_i$, set $(X_{1,i}^\sigma,\ldots,X_{t,i}^\sigma) \sim \eta_i$. Using, sub-additivity of mutual information and independence of $(X_{1,i}^\sigma,\ldots,X_{t,i}^\sigma)$ across the different i's, we infer that

$$\delta n \geq |\Pi(X_1^\sigma,\ldots,X_t^\sigma)| \geq I[X_1^\sigma,\ldots,X_t^\sigma : \Pi(X_1^\sigma,\ldots,X_t^\sigma)] \geq \sum_i I[X_{1,i}^\sigma,\ldots,X_{t,i}^\sigma : \Pi(X_1^\sigma,\ldots,X_t^\sigma)].$$

Averaging over i's and all possible σ's we get,

$$\delta \geq \mathbb{E}_\sigma E_k I[X_{1,k}^\sigma,\ldots,X_{t,k}^\sigma : \Pi(X_1^\sigma,\ldots,X_t^\sigma)] = E_k E_{\sigma_{-k}} E_{\sigma_k} I[X_{1,k}^\sigma,\ldots,X_{t,k}^\sigma : \Pi].$$

Hence, there exists a k and a σ_{-k}, such that

$$\mathbb{E}_{\sigma_k} I[X_{1,k}^\sigma,\ldots,X_{t,k}^\sigma : \Pi] \leq \delta. \tag{12.3.1}$$

We now give a protocol π for computing the multi-party NAND_t function as follows: On inputs $x_1,\ldots,x_t$, the t parties $A_1,\ldots,A_t$ construct n bit inputs $X_1,\ldots,X_t$ for $\mathsf{UDISJ}_{n,t}$ function from $x_1,\ldots,x_t$ as follows: party A_i set the k-bit of X_i to be x_i (i.e., $X_{i,k} = x_i$). For each $i \neq k$, $\sigma_{-k}|_i$ tells which party is active. If $\sigma_{-k}(i) = A_j$, then party A_j sets $X_{j,i}$ with equal probability to 0 or 1, while all other parties ($j' \neq j$) sets $X_{j',i}$ to be 0. Observe, that all of this can be done by the t parties independently using their private randomness and the knowledge of k and σ_{-k}. They, then run the protocol Π on this input $(X_1,\ldots,X_t)$. Since $\mathsf{UDISJ}_{n,t}(X_1,\ldots,X_t) = \mathsf{NAND}_t(x_1,\ldots,x_t)$ and Π is a protocol that computes $\mathsf{UDISJ}_{n,t}$ correctly on every legal input with error at most $1/2-\varepsilon$, we have that π is a protocol that computes NAND_t correctly on every legal input with error at most $1/2-\varepsilon$. Rewriting (12.3.1) in terms of protocol π, we have

$$\frac{1}{t}\cdot\sum_{i=1}^{t} I[Z_i : \pi_{Z_i}] \leq \delta,$$

where Z_i is a random vector defined as follows:

$$Z_i = \begin{cases} \mathbf{0} & \text{with probability } \frac{1}{2} \\ \mathbf{e_i} & \text{the } i\text{th unit vector with probability } \frac{1}{2} \end{cases}$$

We can now complete the lower bound using the properties of the Hellinger distance just as in the case of the disjointness lower bound.

$$\begin{aligned}
\delta &\geq \frac{1}{t} \cdot \sum_{i=1}^{t} I[Z_i : \pi_{Z_i}] \\
&\geq \frac{1}{t} \cdot \sum_{i=1}^{t} h^2(\pi_{\mathbf{0}}, \pi_{\mathbf{e_i}}) && \text{[from Lemma 12.5]} \\
&\geq \frac{1}{t^2} \cdot \left(\sum_{i=1}^{t} h(\pi_{\mathbf{0}}, \pi_{\mathbf{e_i}}) \right)^2 && \text{[By Cauchy-Schwarz]} \\
&= \frac{1}{t^2} \cdot h^2(\pi_{\mathbf{0}}, \pi_{\mathbf{1}}) && \text{[by Claim 12.8]} && (12.3.2) \\
&\geq \frac{1}{2t^2} \cdot \Delta^2(\pi_{\mathbf{0}}, \pi_{\mathbf{1}}) && \text{[by Lemma 12.2]} && (12.3.3) \\
&\geq \frac{2\varepsilon^2}{t^2}
\end{aligned}$$

The only difference from the disjointness proof is Inequality (12.3.2) which is proved in Claim 12.8. This is proved by repeated application of the multi-party Cut-and-Paste Lemma 12.4 and the triangle inequality. The last inequality (12.3.3) follows from the fact that NAND_t takes different answers on inputs $\mathbf{0}$ and $\mathbf{1}$ and hence $\Delta(\pi_{\mathbf{0}}, \pi_{\mathbf{1}}) \geq 2\varepsilon$. This completes the proof of Theorem 12.7. □

This theorem was improved by Gronemeier who proved a $\Omega(n/t)$ lower bound [Gro09].

Claim 12.8. $\sum_{i=1}^{t} h(\pi_{\mathbf{0}}, \pi_{\mathbf{e_i}}) \geq h(\pi_{\mathbf{0}}, \pi_{\mathbf{1}}))$.

Proof. We will illustrate the proof for the case $t = 4$. The general case follows by induction.

$$\begin{aligned}
\text{LHS} &= h(\pi_{0000}, \pi_{1000}) + h(\pi_{0000}, \pi_{0100}) \\
&\quad + h(\pi_{0000}, \pi_{0010}) + h(\pi_{0000}, \pi_{0001}) \\
&\geq h(\pi_{1000}, \pi_{0100}) + h(\pi_{0010}, \pi_{0001}) && \text{[By Triangle inequality]} \\
&= h(\pi_{0000}, \pi_{1100}) + h(\pi_{0000}, \pi_{0011}) && \text{[By Cut-and-Paste Lemma 12.4]} \\
&\geq h(\pi_{1100}, \pi_{0011}) && \text{[By Triangle inequality]} \\
&= h(\pi_{0000}, \pi_{1111}) && \text{[By Cut-and-Paste Lemma 12.4]}
\end{aligned}$$

□

References

[BJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. *An information statistics approach to data stream and communication complexity.* J. Computer and System Sciences, 68(4):702–732, June 2004. (Preliminary Version in *43rd FOCS*, 2002). doi:10.1016/j.jcss.2003.11.006.

[Gro09] Andre Gronemeier. *Asymptotically optimal lower bounds on the NIH-multi-party information complexity of the AND-function and disjointness.* In Susanne Albers and Jean-Yves Marion, eds., *Proc. 26th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 3 of *LIPIcs*, pages 505–516. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2009. doi:10.4230/LIPIcs.STACS.2009.1846.

[KS92] Bala Kalyanasundaram and Georg Schnitger. *The probabilistic communication complexity of set intersection.* SIAM J. Discrete Math., 5(4):545–557, 1992. (Preliminary Version in *2nd Structure in Complexity Theory Conference*, 1987). doi:10.1137/0405044.

[Lin91] Jianhua Lin. *Divergence measures based on the shannon entropy.* IEEE Transaction on Information Theory, 37(1):145–151, jan 1991. doi:10.1109/18.61115.

[Raz92] Alexander A. Razborov. *On the distributional complexity of disjointness.* Theoretical Comp. Science, 106(2):385–390, 1992. doi:10.1016/0304-3975(92)90260-M.

A Weaker proof of Lin's Lemma 12.5

Lemma 12.9. *Let Z be a random variable taking values in $\{z_1, z_2\}$ equally likely and Π a randomized function of Z. Then,*

$$I[Z : \Pi(Z)] \geq \frac{\log_2 e}{2} \cdot h^2(\Pi_{z_1}, \Pi_{z_2}).$$

Proof. Since $x \leq e^x - 1$, we have $\ln y \leq y - 1$. Substituting $y = \sqrt{\frac{p_i}{q_i}}$, we get

$$\frac{1}{2} \cdot \ln \frac{p_i}{q_i} \leq \sqrt{\frac{p_i}{q_i}} - 1.$$

Multiplying by q_i and summing over i, we obtain

$$-\frac{1}{2 \log_2 e} \cdot D(Q \| P) \leq \left(\sum \sqrt{p_i q_i} - 1 \right) = -h^2(Q, P).$$

Using the above bound on divergence, we get

$$\begin{aligned}
I[Z : \Pi_Z] &= \mathbb{E}_{z \leftarrow Z} \left[D(\Pi_z \| \Pi) \right] \\
&= \frac{1}{2} (D(\Pi_{z_1} \| \Pi) + D(\Pi_{z_2} \| \Pi)) \\
&\geq \log e \cdot (h^2(\Pi_{z_1}, \Pi) + h^2(\Pi_{z_2}, \Pi)) \\
&\geq \frac{\log e}{2} \cdot (h(\Pi_{z_1}, \Pi) + h(\Pi_{z_2}, \Pi))^2 && \text{[By Cauchy Schwarz]} \\
&\geq \frac{\log e}{2} \cdot h^2(\Pi_{z_1}, \Pi_{z_2}) && \text{[By triangle inequality]}
\end{aligned}$$

□