Application of Information Theory, Lecture 11

Next-Block Pseudoentropy via KL-hardness

Handout Mode

Iftach Haitner

Tel Aviv University.

January 16, 2020
Section 1

Reminder
Pseudoentropy

Definition 1

\( X \) has \((s, \varepsilon)\)-pseudoentropy at least \( k \), if there exists a random variable \( Y \) with \( H(Y) \geq k \) and \( X \) such that \( Y \) are \((s, \varepsilon)\)-indistinguishable.

\( X \) has \((s, \varepsilon)\)-pseudoentropy at least \( k \) given \( Z \), if there exists a random variable \( Y \) (jointly distributed with \( Z \)) such that \( H(Y|Z) \geq k \) and \((X, Z)\) and \((Y, Z)\) are \((s, \varepsilon)\)-indistinguishable.
Next-block pseudoentropy

Definition 2 (next-block pseudoentropy)

$X = (X_1, \ldots, X_m)$ has $(s, \varepsilon)$ next-block pseudoentropy $k$, if $\exists$ rvs $\{Y_i\}$ (jointly distributed with $X$) such that:

1. $\forall$ s-size $D$ and $i \in [m]: \left| \Pr [D(X_{<i}, X_i) = 1] - \Pr [D(X_{<i}, Y_i) = 1] \right| \leq \varepsilon$

2. $\sum_i H(Y_i | X_{<i}) \geq k$.

Proposition 3

$X = (X_1, \ldots, X_m)$ has $(s, \varepsilon)$ next-block pseudoentropy $k$, iff $X_I$ has pseudoentropy $k/n$ conditioned on $X_{<I}$, for $I \leftarrow [m]$. 
Next-block pseudoentropy from one-way functions

Theorem 4

If \( f : \{0, 1\}^n \mapsto \{0, 1\}^n \) is a \((s, \epsilon)\) OWF, then \( \forall \gamma > 0, g(U_n) = (f(U_n), U_n)\) has \((s', \gamma)\) next-block pseudoentropy \( n + \log(1/\epsilon) - \gamma \), for \( s' = s^{\Omega(1)} / \text{poly}(n, 1/\gamma) \).

For \( \epsilon = 1/n \), \( \gamma = 1 \) and \( s = n^{\log n} \),

\( g \) has \((n^{\Omega(\log n)}, n^{-\Omega(\log n)})\) next-block pseudoentropy \( n + \log n - 1 \).
Section 2

KL-Hardness
KL-hard for sampling

**Definition 5 (KL-hard for sampling)**

Let \((X, B)\) be a rv over \(\mathcal{X} \times \mathcal{B}\). We say that \(B\) is \((s, \delta)\)-KL-hard for sampling given \(X\), if for every \(s\)-size randomized circuit \(S\), it holds that
\[
D(B \parallel S(X) \mid X) > \delta.
\]

Recall

\[
D(X, B \parallel X, S(X)) = D(B \parallel S(x) \mid X) = E_{x \leftarrow X} [D(B \mid x=x \parallel S(X) \mid x=x)]
\]

**Lemma 6**

Let \((X, Z)\) be rv over \(\mathcal{X} \times \{0, 1\}^n\). If \(Z\) is \((s, \delta)\) KL-hard for sampling given \(X\), then for any \(s/O(n)\)-size \(S\):

\[
\sum_{i=1}^{n} D(Z_i \parallel S(Z_{<i}) \mid X, Z_{<i}) > \delta
\]

Equivalently, \(Z_i\) is \((s/O(n), \delta/n)\) KL-hard for sampling given \((X, Z_{<i})\), for \(I \leftarrow [n]\).
Proving Lem 6

Proof: Assume not.

- ∃ s/O(n)-size S such that $\sum_{i=1}^{n} D(Z_i \| S(Z_{<i}) \mid X, Z_{<i}) \leq \delta$
- Consider $S'$ that given $X$ samples $z$ as follows: for $i = 1$ to $n$: $z_i = S(X, z_{<i})$.

\[
D(Z \| S'(X) \mid X) = \sum_{i} D(Z_i \| S'(X)_i \mid X, Z_{<i}) \\
= \sum_{i} D(Z_i \| S(X, Z_{<i}) \mid X, Z_{<i}) \\
\leq \delta.
\]
KL-hardness

- For distribution $P$ over $\mathcal{X} \times \mathcal{Y}$, let $P(y|x) = P_{Y|X}(y|x)$.
- Measure $M: \mathcal{R} \mapsto \mathbb{R}^+$ induces the prob. distribution $C_M(x) = \frac{M(x)}{\sum_b M(b)}$.
- In particular, for $M: \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}^+$, $C_M(y|x) = \frac{M(x,y)}{\sum_b M(x,b)}$.
- For rv $X$ over $\mathcal{X}$, we abuse notation and let $(X, C_M)$ denote the random variable induces by $\Pr[(X, C_M) = (x, y)] = \Pr[X = x] \cdot C_M(y|x)$.

**Definition 7 (KL-estimator)**

Let $(X, B)$ be a rv over $\mathcal{X} \times \mathcal{B}$, and $E: \mathcal{X} \times \mathcal{B} \mapsto \mathbb{R}^+$ be a deterministic function. We say that $P$ is $\delta$-KL estimator of $B$ given $X$ if $D(B\|C_E | X) \leq \delta$.

**Definition 8 (KL-hard)**

$(X, B)$ is $(s, \delta)$-KL hard, if $\exists$ $s$-size $\delta$-KL estimator of $B$ given $X$.

For non trivial $s$, $\delta < \log(\text{Supp}(B)) - H(B|X)$. Indeed, for the constant $E$: $D(B\|C_E | X) = D(B\|U_B | X) = \log(\text{Supp}(B)) - H(B|X)$.

Iftach Haitner (TAU)  
Application of Information Theory, Lecture 11  
January 16, 2020 9 / 24
Lemma 9 (KL-hard for sampling to KL-hardness)

Let \((X, B)\) be rv over \(X \times B\). If \(B\) is \((s, \delta)\)-KL-hard for sampling given \(X\), then \(B\) is \((\Omega(s/|B|), \delta)\)-KL-hard given \(X\).

- Useful for small \(|B|\)
- Other direction also holds.

Proof: HW
Section 3

KL-Hardness to Pseudoentropy
Theorem 10 (KL-hardness to pseudoentropy)

Let \((X, B)\) be rv over \(X \times B\). If \(B\) is \((s, \delta)\)-KL-hard given \(X\), then \(\forall \gamma > 0\), \(B\) has \((s', \gamma)\) pseudoentropy \(H(B|X) + \delta - \gamma\) given \(X\) for
\[s' = s^{\Omega(1) / \text{poly}(\log |X|, \log |B|, 1/\gamma)}.\]

- The key step towards proving Theorem 4
- Other direction also holds.
Generalized distinguishers

A generalized distinguisher $D$ is randomized function that maps strings to $\mathbb{R}^+$. 

$\Delta^D(X, Y) = \mathbb{E}[D(X)] - \mathbb{E}[D(Y)]$

Note that

$$E[D(X, B)] - E[D(X, C)] = \mathbb{E}_X \left[ \sum_a B(a|X) \cdot D(X, a) - C(a|X) \cdot D(X, a) \right]$$

$$= \mathbb{E}_X \left[ \sum_a D(X, a) \cdot (B(a|X) - C(a|X)) \right]$$

For fix $x$, the random variable $2^D$ is defined by

$$2^D(a|x) := \frac{2^{D(x,a)}}{\sum_{b \in B} 2^{D(x,b)}}$$

A conditional version of the Boltzmann distribution.
Distinguishing to KL-divergence

Recall \( 2^D(a|x) := \frac{2^D(x,a)}{\sum_b 2^D(x,b)} \).

Lemma 11

For rv \((X, B)\) and generalized distinguisher \(D\), it holds that

\[
D(B\|2^D \mid X) = H(2^D|X) - H(B|X) - \Delta^D((X, B), (X, 2^D))
\]

- For \(D(x, a) = D(x, a')\), for all \(a, a'\), the distribution \(2^D(a|x)\) is just uniform over \(B\), and the lemma is just \(D(B\|U \mid X) = H(U) - H(B|X)\).
- Assume \(\Delta^D((X, B), (X, 2^D)) \geq 0\), then
  - \(D(B\|2^D \mid X) < H(2^D|X) - H(B|X)\).
  - \(D(B\|2^D \mid X) \geq \delta\) and \(H(2^D|X) \leq H(B|X) + \delta\)
    \(\implies B|X\) has pseudoentropy \(H(2^D|X)\).
Proving Lem 11

Recall $2^D(a|x) := \frac{2^D(x,a)}{\sum_b 2^D(x,b)}$.

Proof:

$$D(B\parallel 2^D \mid X)$$

$$= E_X \left[ \sum_a B(a|X) \log \frac{B(a|X)}{2^D(a|X)} \right]$$

$$= H(2^D|X) - H(B|X) + E_X \left[ \sum_a (B(a|X) - 2^D(a|X)) \log \frac{1}{2^D(a|X)} \right]$$

$$= H(2^D|X) - H(B|X) + E_X \left[ \sum_a (B(a|X) - 2^D(a|X)) \left( \log \left( \sum_b 2^{D(x,b)} \right) - D(x, a) \right) \right]$$

$$= H(2^D|X) - H(B|X) + E_X \left[ \log \left( \sum_b 2^{D(x,b)} \right) (1 - 1) - \sum_a (B(a|X) - 2^D(a|X)) \cdot D(x, a) \right]$$

$$= H(2^D|X) - H(B|X) + E_X \left[ - \sum_a D(x, a)(B(a|X) - 2^D(a|X)) \right]$$

$$= H(2^D|X) - H(B|X) - \Delta^D((X, B), (X, 2^D)).$$
Corollary 12

Let \((X, B)\) and \(2^D\) be as in Lem 12, and let \(C\) be jointly dist. with \(X\) such that 
\[ H(C|X) \geq H(2^D|X). \]
Then \(\Delta^D((X, B), (X, 2^D)) \leq \Delta^D((X, B), (X, C))\).

- \(2^D\) is the closest distribution to \(B|X\), in the eyes of \(D\), among all distributions of conditional entropy \(H(2^D|X)\).

- Among all \(C\)'s with \(\Delta^D((X, B), (X, C)) \leq \Delta^D((X, B), (X, 2^D))\), it holds that \(2^D\) has maximal conditional entropy.

Boltzmann distribution attains maximal entropy under a linear restriction.
Proving corollary 12

Proof: Let $C$ be such that $H(C|X) \geq H(2^D|X)$.

$$D(C\|2^D \mid X) = H(2^D|X) - H(C|X) - \Delta^D((X, C), (X, 2^D))$$

$$\leq -\Delta^D((X, C), (X, 2^D))$$

$$= \Delta^D((X, B), (X, C)) - \Delta^D((X, B), (X, 2^D))$$

Hence, $\Delta^D((X, B), (X, C)) \geq \Delta^D((X, B), (X, 2^D))$. 
Universal distinguishing to small KL-divergence

**Lemma 13**

Let \((X, B)\) be rv with \(H(B|X) \leq \log |B| - \delta\), and let \(D\) be gen. distinguisher with \(\Delta^D((X, B), (X, C)) > \gamma\) for all \(C\) with \(H(C|X) \geq H(B|X) + \delta\). Then \(\exists k \in [0, (\log |B|)/\gamma]\) such that \(D(B\|2^{kD} | X) \leq \delta\).

Hence, universal \(D\) that distinguishes \(B\) from all high entropy \(C\), can be used to approximate \(B\) to within small KL divergence.

**Proof:** Let \(q = |B|\) and \(k_0 = (\log q)/\gamma\). By Lem 11,

\[
\Delta^D((X, B), (X, 2^{k_0 D})) = \frac{1}{k_0} \left( H(2^{k_0 D} | X) - H(B|X) - D(B\|2^{k_0 D} | X) \right) \leq (\log q)/k_0 = \gamma
\]

1. \(H(2^{k_0 D} | X) < H(B|X) + \delta\) (by assumption about \(D\))
2. \(H(2^{0} | X) = \log q \geq H(B|X) + \delta\) (by assumption about \((X, B)\))
3. \(H(2^{kD} | X)\) is continues function of \(k\)

Hence, \(\exists k \in [0, k_0]\) such that \(H(2^{kD} | X) = H(B|X) + \delta\). Thus by Lem 11

\[
D(X, B\|X, 2^{kD}) = H(2^{kD} | X) - H(B|X) - \Delta^{kD}((X, B), (X, 2^{kD}))
= \delta - k \cdot \Delta^D((X, B), (X, 2^{kD})) \leq \delta - k\gamma < \delta.
\]
**MiniMax theorem**

Let $n, m \in \mathbb{N}$ and let $M \in \mathbb{R}^{[n] \times [m]}$.

- Let $\zeta_t$ be the set of all distributions over $[t]$
- $\maxmin(M) = \max_{X \in \zeta_n} \min_{Y \in [m]} \mathbb{E}_{X \leftarrow X} [M(x, y)]$
- $\minmax(M) = \min_{Y \in \zeta_m} \max_{X \in [n]} \mathbb{E}_{Y \leftarrow Y} [M(x, y)]$
- Clearly, $\maxmin(M) \leq \minmax(M)$

**Theorem 14 (Von Neumann)**

$maxmin(M) = minmax(M)$

- Example: algorithm that preform well on any distribution of problems, implies a randomized algorithm that solves well any problem
- Can be proved via Duality of Linear Programs
Proving Theorem 10

**Theorem 15 (KL-hardness to pseudoentropy, restatement of Theorem 10)**

Let \((X, B)\) be rv over \(\mathcal{X} \times \mathcal{B}\). If \(B\) is \((s, \delta)\)-KL-hard given \(X\), then \(\forall \gamma > 0\), \(B\) has \((s', \gamma)\) pseudoentropy \(H(B|X) + \delta - \gamma\) given \(X\) for \(s' = s^{\Omega(1)} / \text{poly}(\log |\mathcal{X}|, \log |\mathcal{B}|, 1/\gamma)\).

**Proof:** Assume \(\forall C \in \mathcal{C} = \{C : H(C|X) \geq H(B|X) + \delta - \gamma\}\), exits a \(s'\)-size \(D\) with \(\Delta^D((X, B), (X, C)) > \gamma\).

- Two-player game: Player \(P_1\) picks \(C \in \mathcal{C}\). Player \(P_2\) picks \(s'\)-size \(D\). Payoff is \(\Delta^D((X, B), (X, C))\).
- For any convex combination \(C\) of \(\mathcal{C}\), it holds that \(C \in \mathcal{C}\)
- \(P_1\) has no mix strategy to achieve payoff \(\leq \gamma\).
- By MiniMax, \(P_2\) has mix strategy \(D\) to achieve payoff \(> \gamma\) for any \(C \in \mathcal{C}\)
- By Lem 13, \(\exists k \in [0, (\log |\mathcal{B}|)/\gamma]\) such that \(D(B\|2^{kD} | X) \leq \delta - \gamma\).
- But what is the size of \(D\)!!
Efficient variant of $\mathcal{D}$

- $\tilde{\mathcal{D}}(x, b)$ samples $\nu = O((\log |\mathcal{X}| + \log |\mathcal{B}|)/\gamma^2)$ samples from $\mathcal{D}$ and returns their average prediction on $(x, b)$.

- By Hoeffding: $\forall (x, b) \in \mathcal{X} \times \mathcal{B}$: $|\tilde{\mathcal{D}}(x, b) - \mathcal{D}(x, b)| \leq \gamma/2$

- $\exists$ deterministic circuit $\hat{\mathcal{D}}$ of size $\nu \cdot s$ for which the above holds.

- Hence, $\Delta^{\hat{\mathcal{D}}}(\mathcal{X}, \mathcal{B}) > \gamma/2$ for all $C \in \mathcal{C}$

- By Lem 13, $\exists k \in [0, (\log |\mathcal{B}|)/\gamma]$ such that $D(B\|2^k\hat{\mathcal{D}} \mid X) \leq \delta - \gamma/2$.

- Exists (?) circuit $P$ of size $\text{poly}(s', \log |\mathcal{X}|, \log |\mathcal{B}|, 1/\gamma)$ approximating $2^k\hat{\mathcal{D}}$ so that

  $$D(B\|P \mid X) \leq D(B\|2^k\hat{\mathcal{D}} \mid X) + \gamma/2 \leq \delta$$
Section 4

From One-Way Functions to Next-Block Pseudoentropy
Lemma 16

Let $f : \{0, 1\}^n \rightarrow \{0, 1\}^n$ is a $(s, \varepsilon)$ OWF, then $U_n$ is $(s, \log(1/\varepsilon))$ KL-hard for sampling given $f(U_n)$.

Proof: Assume not.

- Let $S$ be $s$-size circuit s.t $D(f(U_n)\|S(f(U_n)) \mid U_n) \leq \log(1/\varepsilon)$
- Let $g(y, x) := (y == f(x))$.
- By date processing, $D(g(f(U_n), U_n)\|g(f(U_n), S(f(U_n)))) \leq \log(1/\varepsilon)$.
- But $D(g(f(U_n), U_n)\|g(f(U_n), S(f(U_n)))) > D(1\|\varepsilon) = \log(1/\varepsilon)$
Next-block pseudoentropy from OWfs, proving Theorem 4

Theorem 17 (restating Theorem 4)

If \( f : \{0, 1\}^n \mapsto \{0, 1\}^n \) is a \((s, \varepsilon)\) OWF, then \( \forall \gamma > 0, g(U_n) = (f(U_n), U_n) \) has \((s', \gamma)\) next-block pseudoentropy \( n + \log(1/\varepsilon) - \gamma \), for \( s' = s^{\Omega(1)}/\text{poly}(n, 1/\gamma) \).

Proof: Let \( Z = U_n, X = f(Z), I \leftarrow [n] \) and \( \gamma > 0 \).

1. By Lem 16 and Lem 6: \( Z_I \) is \((s' = s/\text{poly}(n), \varepsilon' = \log(1/\varepsilon)/n)\) KL-hard for sampling given \((X, Z_{<I})\).

2. By Lem 9: \( Z_I \) is \((s', \varepsilon')\) KL-hard given \((X, Z_{<I})\).

3. By Thm 10: \( Z_I \) has \((s'', \gamma)\) pseudoentropy \( H(Z_I|X, Z_{<I}) + \varepsilon' - \gamma \) given \((X, Z_{<I})\), for \( s'' = s^{\Omega(1)}/\text{poly}(n, 1/\gamma) \).

4. Let \( T = (X, Z) \) and \( J \leftarrow [2n] \).

5. By (3), \( T_J \) has \((s'', \gamma)\) pseudoentropy \( H(T_J|T_{<J}) + (\varepsilon' - \gamma)/2 \) given \( T_{<J} \).

6. By Prop 3, \( T \) has next-block pseudoentropy \( 2n(H(T_J|T_{<J}) + (\varepsilon' - \gamma)/2) \implies T \) has next-block pseudoentropy \( n + n(\varepsilon' - \gamma) = n + \log(1/\varepsilon) - n\gamma \).