Application of Information Theory, Lecture 3
Graph Covering, Differential Entropy

Handout Mode

Iftach Haitner

Tel Aviv University.

March 22, 2018
Part I

Applications to Graph Covering
Graph Covering

- How many graphs of certain type it takes to cover the full graph?
- \( K_n \) — the complete graph over \([n]\)
- Let \( G_1, \ldots, G_t \) be bipartite graphs over \([n]\) with \( \bigcup_i G_i = K_n \).
  What can we say about \( t \)?
- Clearly, \( t \geq \frac{(n/2)^2}{n/2} \approx 2 \), but can we give a better bound?

**Theorem 1**

Let \( G_1, \ldots, G_t \) be bipartite graphs over \([n]\) with \( \bigcup_{i=1}^t G_i = K_n \), then \( t \geq \log n \).

**Proof:** Let \( \chi(G) \) be the chromatic number of \( G \).

- \( \chi(G_i) \leq 2 \) and \( \chi(K_n) = n \).
- \( \chi(G \cup G') \leq \chi(G) \cdot \chi(G') \).

\[
\Rightarrow \chi(\bigcup_{i=1}^t G_i) \leq 2^t \\
\Rightarrow t \geq \log n
\]
Proving Thm 1 using entropy

- $G_i = (A_i, B_i, E_i)$
- $X \leftarrow [n]$
- $Y_i = \begin{cases} 0, & X \in A_i \\ 1, & X \in B_i \end{cases}$
- $X$ is determined by $Y_1, \ldots, Y_t$ (?)

\[
0 = H(X|Y_1, \ldots, Y_t) = H(X, Y_1, \ldots, Y_t) - H(Y_1, \ldots, Y_t) \\
\geq H(X) - \sum_i H(Y_i) \\
\geq \log n - t.
\]
Extensions

- \text{nonIs}(G) — non-isolated vertices in \( G \).

\textbf{Theorem 2}

Let \( G_1, \ldots, G_t \) be bipartite graphs over \([n]\) with \( \bigcup_{i=1}^{t} G_i = K_n \), then
\[
\frac{1}{n} \sum_{i=1}^{t} |\text{nonIs}(G_i)| \geq \log n.
\]

\textbf{Definition 3 (graph content)}

Let \( G \) be a graph over \([n]\), let \( Z \leftarrow \text{nonIs}(G) \) and let \( \hat{\chi} \) be a (valid) coloring of \( G \) such that \( H(\hat{\chi}(Z)) \) is minimal. Then \( \text{content}(G) := \frac{|\text{nonIs}(G)|}{n} \cdot H(\hat{\chi}(Z)) \).

\textbf{Theorem 4}

Let \( G_1, \ldots, G_t \) be graphs over \([n]\) with \( \bigcup_{i=1}^{t} G_i = K_n \). Then
\[
\sum_{i} \text{content}(G_i) \geq \log n.
\]

- Since \( \text{content}(G) \leq \frac{|\text{nonIs}(G)|}{n} \) for bipartite \( G \), Thm 4 yields Thm 2.
Proving Thm 4

- Let $\chi_i$ be a (valid) coloring of $G_i$.

- Let $X \leftarrow [n]$, and let
  
  $$Y_i = \begin{cases} 
  \chi_i(X) & X \in \text{nonls}(G_i) \\
  \chi_i(Z_i) & \text{otherwise, for } Z_i \leftarrow \text{nonls}(G_i) \text{ (ind. of the other } Z \text{'s)}.
  \end{cases}$$

- $X$ is determined by $Y_1, \ldots, Y_t$ (?)
  
  $$0 = H(X|Y_1, \ldots, Y_t) = H(X, Y_1, \ldots, Y_t) - H(Y_1, \ldots, Y_t)$$
  
  $$\geq H(X) + H(Y_1, \ldots, Y_t|X) - \sum_i H(Y_i)$$
  
  $$= \log n + H(Y_1, \ldots, Y_t|X) - \sum_i H(Y_i).$$

- $Y_1, \ldots, Y_t$ are independent conditioned on $X$ —
  
  $$\Pr [Y_1 = y_1 \land Y_2 = y_2 | X = x] = \Pr [Y_1 = y_1 | X = x] \cdot \Pr [Y_2 = y_2 | X = x]$$

- Hence, $H(Y_1, \ldots, Y_t|X) = \sum_i H(Y_i|X)$ (board)

- We conclude that $\sum_i H(Y_i) - \sum_i H(Y_i|X) \geq \log n$

- Since $H(Y_i) = H(\chi_i(Z_i))$ and $H(Y_i|X) = (1 - \frac{|\text{nonls}(G_i)|}{n}) \cdot H(\chi_i(Z_i))$, it follows that $\sum_i H(\chi_i(Z_i)) \frac{|\text{nonls}(G_i)|}{n} \geq \log n$. □
Extension

Let $\alpha(G)$ be the size of the maximal independent set in $G$.

**Theorem 5**

Let $G, G_1, \ldots, G_t$ be graphs over $[n]$ with $\bigcup_{i=1}^{t} G_i = G$, then

$$\sum \text{content}(G_i) \geq \log \frac{n}{\alpha(G)}.$$ 

**Proof:** HW
Scrambling permutations

**Theorem 6**

Let $S$ be a set of permutations over $[n]$ s.t. for any triplet $(i, j, k)$ of distinct elements of $[n]$, exists $\pi \in S$ with $\pi(i) < \pi(j) < \pi(k)$ or $\pi(i) > \pi(j) > \pi(k)$. Then $|S| \geq 2^\frac{2}{\log e} \log n$.

- For $\pi \in S$, the graph $G_\pi = (V, E_\pi)$ is defined by:
  - $V = \{(i, j) \in [n]^2 : i \neq j\}$
  - $E_\pi = \{((i, j), (k, j)) \in V^2 : \pi(i) < \pi(j) < \pi(k) \lor \pi(i) > \pi(j) > \pi(k)\}$

- $G = \bigcup_{\pi \in S} G_\pi$ has $n$ connected components, each consists of $(n-1)$-vertex cliques: $C^j = \{(i, j) : i \in [n] \setminus \{j\}\}$ for each $j \in [n]$.

- $G_\pi$ consists of $n$ complete bipartite graphs (two are empty): $G^i_\pi = \{(i, j) : \pi(i) \leq \pi(j)\}$ and $\{(i, j) : \pi(i) > \pi(j)\}$ for each $j \in [n]$.

- $\sum_{\pi} \sum_i \text{content}_{C_i}(G^i_\pi) = \sum_{\pi} \sum_i h(|\{(i, j) : \pi(i) \leq \pi(j)\}| / (n-1))$

- $\sum_i |S| \cdot h(\frac{i}{n-1}) = |S| \cdot \sum_i h(\frac{i}{n-1}) \leq |S| (n-1) \cdot \frac{\log e}{2}$

- By Thm 4: $\sum_{\pi} \text{content}_{C_i}(G^i_\pi) \geq \log(n-1)$

- Hence, $|S| (n-1) \cdot \frac{\log e}{2} \geq n \cdot \log(n-1)$, and the proof follows. □
Part II

Differential Entropy
Entropy of continuous random variable

- Entropy of discrete random variable: $H(X) = - \sum_i p_i \log p_i$
- Also used when $X$ has infinite support (entropy might be infinite)
- Continues random variable is defined by its density function: $f: \mathbb{R} \mapsto \mathbb{R}^+$, for which $\int_{\mathbb{R}} f(x) dx = 1$.
- $F_X(x) := \Pr [X \leq x] = \int_{-\infty}^{x} f(x) dx$
- $E X = \int x \cdot f(x) dx$ and $V X = \int x^2 \cdot f(x) dx - (E X)^2$
- Examples: $X \sim [0, 1], X \sim N(0, 1)$
- $H(X)$ must be infinite! It takes infinite number of bits to describe $X$
- The differential entropy of $X$ is defined by $h(X) = - \int f(x) \log f(x) dx$.
- We focus on cases where $h(X)$ is well defined.
- Since $h$ is a function of the density function, we sometimes write $h(f)$
- If not stated otherwise, we integrate over $\mathbb{R}$
Intuition for definition of $h$

Let $X^{\Delta}$ be rounding of $X$ for precision $\Delta$:

$X^{\Delta} \sim (\ldots, p_{-2}, p_{-1}, p_0, p_1, p_2, \ldots)$,

where $p_i = \int_{i \cdot \Delta}^{(i+1) \cdot \Delta} f(x) \, dx = f(x_i) \cdot \Delta$

for some $x_i \in [i \cdot \Delta, (i + 1) \cdot \Delta]$ (?).

$H(X^{\Delta}) = - \sum_{i=-\infty}^{\infty} p_i \log p_i$

$$H(X^{\Delta}) = - \sum_{i=-\infty}^{\infty} f(x_i) \cdot \Delta \cdot \log(f(x_i) \cdot \Delta) = \sum_{i=-\infty}^{\infty} f(x_i) \cdot \Delta \cdot (\log f(x_i) + \log \Delta)$$

$$= - \sum_{i=-\infty}^{\infty} f(x_i) \cdot \log f(x_i) \cdot \Delta - \left( \sum_{i=-\infty}^{\infty} f(x_i) \cdot \Delta \right) \log \Delta$$

$\lim_{\Delta \to 0} H(X^{\Delta}) = h(X) - \lim_{\Delta \to 0} \log \Delta$

Hence, $\lim_{\Delta \to 0} (H(X^{\Delta}) + \log \Delta) = h(x)$

Intuitively, $h(X)$ is the entropy of $X$ plus const $(\lim_{\Delta \to 0} - \log \Delta)$.

Note that $\lim_{\Delta \to 0} - \log \Delta = \infty$
Properties of the entropy function

\[ h(X) = - \int f(x) \log f(x) \, dx \]

- Shift invariant: \( h(f) = h(g) \) for \( g(x) = f(x + a) \)
- \( h(f) \) might be infinite
  
  For any discrete \( X \) exists \( f \) with \( h(f) = H(X) \).
- \( h(X) \) might be negative
- Example: \( X \sim [0, a] \) – \( f(x) = \frac{1}{a} \) on \([1, a]\)
  
  \[ -\int f(x) \log f(x) \, dx = -\log \frac{1}{a} = \log a. \] Negative for \( a < 1 \).
- \( h(X) \) should be interpreted as the uncertainty up to a certain constant
- Used for comparing two distributions
Common distribution (in nature)

- The uniform distribution: $X \sim [a, b]$
- Normal (Gaussian) distribution: (we focus on $E = 0$ and $V = 1$)
  $$X \sim \mathcal{N}(0, 1): \quad f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}$$
- Boltzmann (Gibbs) distribution:
  $$X \in \{E_1, E_2, \ldots, E_m\}, \quad \Pr[X = E_i] = C \cdot e^{-\beta E_i} \text{ for } \beta > 0 \text{ (the distribution constant)}$$
  and $C = 1/\sum_i e^{-\beta E_i}$.
  
  - Describes a (discrete) physical system that can take states $\{1, \ldots, m\}$ with energies $E_1, \ldots, E_m$.
  - Probability is inverse to the energy

- Why are these distributions so common?
- What is common to these distributions?
Second law of thermodynamics

- The entropy of a closed physical system never decreases.
- If we wait enough time, the system tends to be in maximal entropy.
- If there are constrains, the it tends to be in maximal entropy under this constrains.
- This suggests that distributions that are common in nature, are distributions of maximal entropy, under some constrains.
The normal distribution

- \( X \sim N(0, 1) \): 
  \[
f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}
\]

- Why is it so common?
- Answer: the central limit theorem (CLT):
  
  Let \( X_1, \ldots, X_n \) be iid with \( \mathbb{E} X_i = 0 \) and \( \mathbb{V} X_i = 1 \). Then
  \[
  \lim_{n \to \infty} \frac{\sum_i X_i}{\sqrt{n}} = N(0, 1).
  \]

- But why does it converge to \( N(0, 1) \)?
- CLT holds also in many other variants: not id, not fully independent, ...

- We know that \( \mathbb{E} \frac{\sum_i X_i}{\sqrt{n}} = 0 \) and \( \mathbb{V} \frac{\sum_i X_i}{\sqrt{n}} = 1 \), but it could have converge to any other distribution with these constraints.

- The reason is that \( N(0, 1) \) has the highest entropy among all distribution with \( \mathbb{E} = 0 \) and \( \mathbb{V} = 1 \).

- CLT and the normal distribution where known and studied way before Shannon, yet this striking property was not known until his theory.
The normal distribution, cont.

**Theorem 7**

\[ h(X) \leq h(N(0, 1)), \text{ for any rv } X \text{ with } \text{Var } X = 1. \]

- Among the distributions of \( \text{Var } = 1 \), the distribution \( N(0, 1) \) has **maximal** entropy.
- Generalizes to any variance:
  \[ h(X) \leq h(N(0, \text{Var } X)) = \frac{1}{2} \log(2\pi e) \cdot \text{Var } X \]

Let \( g \) be a density function with \( \int g(x)x^2dx = 1 \), and let \( f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \).
We will show that

1. \( -\int g(x) \log g(x)dx \leq -\int g(x) \log f(x)dx \)
2. \( -\int g(x) \log f(x)dx = -\int f(x) \log f(x)dx \)
\[- \int g(x) \log g(x) \, dx \leq - \int g(x) \log f(x) \, dx \]

**Claim 8**

\[- \int g(x) \log g(x) \, dx \leq - \int g(x) \log q(x) \, dx \] for any two density functions \( q, g \).

**Proof:**

- **Jensen:** For any function \( t \) and density function \( \lambda \):
  \[
  \int \lambda(x) \log t(x) \, dx \leq \log \int \lambda(x) t(x) \, dx
  \]

- Assume for simplicity that \( g(x) > 0 \) for all \( x \).

- By Jensen,
  \[
  \int g(x) \log \frac{q(x)}{g(x)} \, dx \leq \log \int g(x) \frac{q(x)}{g(x)} \, dx = \log 1 = 0
  \]

- Hence,
  \[
  - \int g(x) \log g(x) \, dx \leq - \int g(x) \log q(x) \, dx
  \]
\[- \int g(x) \log f(x) \, dx = - \int f(x) \log f(x) \, dx\]

**Claim 9**

Exists $c \in \mathbb{R}$ such that $- \int g(x) \log f(x) \, dx = c$ for any density function $g$ with $\int g(x) x^2 \, dx = 1$.

Hence, $- \int g(x) \log f(x) \, dx = - \int f(x) \log f(x) \, dx$.

Proof:

\[- \int g(x) \log f(x) \, dx = - \int g(x) \log \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \, dx\]

\[= - \int g(x) \left( \log \frac{1}{\sqrt{2\pi}} - \frac{x^2}{2} \cdot \log e \right) \, dx\]

\[= - \log \frac{1}{\sqrt{2\pi}} \int g(x) \, dx + \frac{\log e}{2} \int g(x) x^2 \, dx\]

\[= - \log \frac{1}{\sqrt{2\pi}} + \frac{\log e}{2}.\]
The Boltzmann distribution

- States \( \{1, \ldots, m\} \), energies \( E_1, \ldots, E_m \).
- \( \Pr[X = E_i] = C \cdot e^{-\beta E_i} \) for \( \beta > 0 \) and \( C = 1 / \sum_i e^{-\beta E_i} \)
- We will denote it by \( \sim B(\beta, E_1, \ldots, E_m) \)
- Like the exponential distribution (i.e., \( f(x) = \lambda e^{-\lambda x} \)), but discrete.
  - Describes a (discrete) physical system that can take states \( \{1, \ldots, m\} \) with energies \( E_1, \ldots, E_m \).
  - Probability is inverse to energy

**Theorem 10**

Let \( X \sim B(\beta, E_1, \ldots, E_m) \). Then \( H(Y) \leq H(X) \) for any rv \( Y \) over \( \{E_1, \ldots, E_m\} \), with \( E_Y = E_X \).

- The Boltzmann distribution is maximal among all distributions of the same energy.
Proving Theorem 10

- \( B(\beta, E_1, \ldots, E_m) \) and \( E Y = E X \)
- Let \( X \sim (p_1, \ldots, p_m) \) and \( Y \sim (q_1, \ldots, q_m) \) over \( \{E_1, \ldots, E_m\} \).
- \( H(Y) \leq \sum_i q_i \log p_i \) \hspace{1cm} (Q3 in Handout 1)
- Let \( C = 1/\sum_i e^{-\beta E_i} \).

Then
\[
\sum_i q_i \log p_i = \sum_i q_i \log (C \cdot e^{-\beta E_i})
\]
\[
= \sum_i q_i \log C - \sum_i q_i \cdot \beta E_i \cdot \log e
\]
\[
= \log C - \beta \cdot \log e \cdot \sum_i q_i E_i
\]
\[
= \log C - \beta \cdot \log e \cdot E X
\]

- Hence, \( \sum_i q_i \log p_i = \sum_i p_i \log p_i \). \( \Box \)
The uniform distribution

- \( X \sim [a, b] \).
- \( E X = \frac{1}{2} (a + b) \) and \( V X = \frac{1}{12} (b - a)^2 \)
- What come to mind when saying “\( X \) takes values in \([0, 1]\)”.  

**Theorem 11**

\[ h(X) \leq -h(\sim [a, b]) \], for any RV with \( \text{Supp}(X) \subseteq [a, b] \).

Proof: HW
Using diff. entropy to bound discrete entropy

**Proposition 12**

Let $X \sim (p_1, p_2, \ldots)$, then $H(X) \leq \frac{\log 2\pi e}{2} \cdot \left( \sum_{i=1}^{\infty} p_i \cdot i^2 - \left( \sum_{i=1}^{\infty} p_i \cdot i \right)^2 + \frac{1}{12} \right)$

We assume w.l.o.g. that $p_i = \Pr[X = i]$.

- Let $U \sim [0, 1]$, let $\tilde{X} = X + U$ and let $f_{\tilde{X}}$ be the density function of $\tilde{X}$.

\[
H(X) = - \sum_{i=1}^{\infty} p_i \log p_i
\]

\[
= - \sum_{i=1}^{\infty} \left( \int_{i}^{i+1} f_{\tilde{X}}(x) dx \right) \cdot \log p_i = - \sum_{i=1}^{\infty} \int_{i}^{i+1} f_{\tilde{X}}(x) \log p_i dx
\]

\[
= - \sum_{i=1}^{\infty} \int_{i}^{i+1} f_{\tilde{X}}(x) \log f_{\tilde{X}}(x) dx \quad (f_{\tilde{X}}(x) = p_i \text{ for all } x \in [i, i+1])
\]

\[
= - \int_{1}^{\infty} f_{\tilde{X}}(x) \log f_{\tilde{X}}(x) dx
\]

\[
= h(\tilde{X})
\]
Using diff. entropy to bound discrete entropy, cont.

- Hence,

\[ H(X) = h(\tilde{X}) \leq \frac{1}{2} \log(2\pi e) V(\tilde{X}) = \frac{1}{2} \log(2\pi e) (V(X) + V(U)) = \frac{\log 2\pi e}{2} \cdot \left( \sum_{i=1}^{\infty} p_i \cdot i^2 - \left( \sum_{i=1}^{\infty} p_i \cdot i \right)^2 \right) + \frac{1}{12} \]

- How good is this bound?

- Let \( X \sim (\frac{1}{2}, \frac{1}{2}) \). Hence, \( V[X] = \frac{1}{4} \) and \( H(X) = 1 \).

- Proposition 12 grants that \( H(X) \leq \frac{\log 2\pi e}{2} \left( \frac{1}{4} + \frac{1}{12} \right) \sim 1.255 \)