PCMI GSS Asymptotic Enumeration: Problem Set 4

Lecturer: Jinyoung Park; TA: Bob Krueger

July 11, 2025

The $\star$'s roughly indicate difficulty. Hints are given as footnotes. Please find a group to work with and don't be afraid to ask questions! Also please continue working on problems from previous problem sets (attached at the end of this document).

1. $\star\star$ Prove the following using entropy.[1]

   Proposition 1. *Let G be a d-regular bipartite graph on n vertices. For any constant q,*
   $$\log c_q(G) \le \frac{n}{2} \cdot \left(\log\left(\left\lfloor\frac{q}{2}\right\rfloor \left\lceil\frac{q}{2}\right\rceil\right) + O(1/d)\right).$$

2. $\star\star$ Prove the following lemma using one of the following approaches. (If you have a different proof, please tell Bob!)

   Lemma 2. *Let G be a graph of maximum degree d, and let $v \in V(G)$. Let $\mathcal{A}_v(k) = \{A \subseteq V(G) : v \in A, |A| = k, G[A]$ is connected$\}$. Then $|\mathcal{A}_v(k)| \le (ed)^k$.*

   (a) Let $X = V(G)_p$ be a random subset of $V(G)$ where each vertex is included independently with probability p. Estimate the probability that the component of $G[X]$ containing v has k vertices, and optimize p.[2]

   (b) Let
   $$T_{G,v}(x) = \sum_{\text{tree } T\subseteq G: v\in V(T)} x^{e(T)}.$$
   Show that
   $$T_{G,v}(x) \le \prod_{w\in N(v)} \left(1 + xT_{G-v,w}(x)\right),$$
   and establish that $T_{G,v}(\frac{1}{ed}) \le e$.[3]

[1] Recall the proof of $i(G) \le \frac{n}{2}(1 + O(1/d))$ using entropy. The heart of the matter is showing that $H(f_x|f(N_x)) + \frac{1}{d}H(f_{N_x}|f(N_x)) \le \log\left(\left\lfloor\frac{q}{2}\right\rfloor \left\lceil\frac{q}{2}\right\rceil\right)$ for a vertex x. If $f(N_x) = C$, then the first term is at most $\log(q - |C|)$ and the second term is at most $\log |C|$, and thus sum of these is at most $\log\left(\left\lfloor\frac{q}{2}\right\rfloor \left\lceil\frac{q}{2}\right\rceil\right)$.

[2] This probability is $\sum_{A\in\mathcal{A}_v(k)} p^k(1-p)^{|N(A)\setminus A|}$. Use that this probability is at most 1 and $|N(A) \setminus A| \le d|A|$.

[3] Use induction and observe that the coefficients of $T_{G,v}(x)$ upper bound what we want to count.

(c) Show that the number of k-vertex rooted subtrees of the infinite rooted d-ary tree is exactly $\frac{1}{k}\binom{dk}{k-1}$.[4]

[4] Bob doesn't know a simple way of proving this.

Problem Set 3

1. $\star$ Recall of the definition of a ψ-approximation: for $A \subseteq \mathcal{O}$, a ψ-approximation $(S, F) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$ is such that $[A] \subseteq S$ ($[A] = \{v : N(v) \subseteq N(A)\}$ is the closure of A), $F \subseteq N(A)$, $d_F(u) \geq d - \psi$ for all $u \in S$, and $d_S(v) \leq \psi$ for all $v \in \mathcal{E} \setminus F$. Show that[5]

$$|S| \leq |F| + \frac{t\psi}{d - \psi},$$

where $t = |N(A)| - |[A]|$.

2. $\star\star$ The goal of this problem is to derive $i(Q_d) \leq (1 + o(1))2\sqrt{e}2^{2^{d-1}}$. Recall that $N := |V(Q_d)| = 2^d$. We will use the three lemmas below for the proof.

Lemma 3. *In Q_d, let $\mathcal{G}(a, g)$ be the set of all 2-linked $A \subseteq \mathcal{O}$ such that $|[A]| = a$, $|N(A)| = g$. Then there exists a constant $c > 0$ such that*

$$|\mathcal{G}(a, g)| \leq 2^{d-1} \cdot 2^{g - c(g-a)}.$$

Below are well-known isoperimetric inequalities for Q_d.

Lemma 4. *For $A \subseteq \mathcal{O}$,*

(a) If $|A| \leq d/10$, then $|N(A)| \geq d|A| - |A|^2$.
(b) If $|A| \leq d^{10}$, then $|N(A)| \geq d|A|/10$.
(c) If $|A| \leq 2^{d-2}$, then $|N(A)| \geq (1 + \Omega(1/\sqrt{d}))|A|$.

The next lemma is useful for counting connected subsets.

Lemma 5. *Let G be a graph of maximum degree at most d, and let $v \in V(G)$. Let $\mathcal{A}_v(k) = \{A \subseteq V(G) : v \in A, |A| = k, G[A]$ is connected$\}$. Then $|\mathcal{A}_v(k)| \leq (ed)^k$.*

Now, derive $i(Q_d) \leq (1 + o(1))2\sqrt{e}2^{2^{d-1}}$ following the given steps.

(a) Prove that if I is an independent set of Q_d, then either $|I \cap \mathcal{E}|$ or $|I \cap \mathcal{O}|$ is at most $N/4 = 2^{d-2}$. Use this to show that

$$i(Q_d) \leq 2 \cdot 2^{N/2} \sum_{\substack{A \subseteq \mathcal{E} \\ |[A]| \leq N/4}} 2^{-|N(X)|}.$$

[5]Upper bound $e(S, F) + e(S, N(A) \setminus F) + e(S, \mathcal{E} \setminus N(A))$.

(b) Using (a), show that[6]

$$i(Q_d) \le 2 \cdot 2^{N/2} \exp\left(\sum_{\substack{A \subseteq \mathcal{E} \\ A \text{ is 2-linked} \\ 1 \le |[A]| \le N/4}} 2^{-|N(A)|} \right)$$

(c) Compute the sum of the terms with $|A| = 1$.

(d) Use Lemma 4 and Lemma 5 to show that the sum of the terms with $2 \le |A| \le d/10$ is $o(1)$.

(e) Use Lemma 3 and Lemma 4 to show that the sum of the remaining terms is $o(1)$.

[6] Break A up into its '2-linked components' $A_1, \dots, A_k$, and use that $N(A)$ is the disjoint union of $N(A_1), \dots, N(A_k)$.

Problem Set 2

1. $\star - \star\star$ Try to come up with asymptotically accurate lower bounds construction for the following counts. Convince yourself why your construction should be asymptotically best possible.

   (a) The number of 4-colorings of Q_d.[7]

   (b) The number of 5-colorings of Q_d.[8]

   (c) The number of (rooted) graph homomorphisms from Q_d to $\mathbb{Z}$. That is, $f : V(Q_d) \to \mathbb{Z}$ with $f(0) = 0$ and $|f(x) - f(y)| = 1$ for every edge xy of Q_d.[9]

   (d) The number of (rooted) '3-Lipschitz functions' on Q_d, which are functions $f : V(Q_d) \to \mathbb{Z}$ with $f(0) = 0$ and $|f(x) - f(y)| \leq 3$ for every edge xy of Q_d.[10]

2. $\star$ Recall that $i(Q_d) = (1+o(1))2\sqrt{e}2^{2^{d-1}}$. Compute the $o(1)$ error term to order 2^{-d} by considering 'defects' consisting of 1 or 2 nearby vertices.[11] If you're feeling brave, you can compute it to order 2^{-2d} by considering 'defects' of up to 3 nearby vertices. (Be careful not to overcount. There is a systematic/algorithmic way, known as the cluster expansion method from statistical physics, for computing these finer asymptotics.)

3. $\star\star$ In this exercise, you will prove the following container lemma from lecture (without assuming G is bipartite):

   Lemma 6. *Let G be an n-vertex d-regular graph. For every $\varepsilon > 0$, there exists $\mathcal{C} \subseteq 2^{V(G)}$ such that*

   - *for every independent set I of G, there exists $C \in \mathcal{C}$ such that $I \subseteq C$,*
   - *$|\mathcal{C}| \leq \binom{n}{\leq n/\varepsilon d}$, and*
   - *for every $C \in \mathcal{C}$, $|C| \leq \frac{n}{\varepsilon d} + \frac{n}{2-\varepsilon}$.*

   Recall that $\Delta(G)$ is the maximum degree of G, and $G[U]$ denotes the subgraph of G induced by U.

   (a) Let $C \subseteq V(G)$ and $\varepsilon > 0$. Show that if $|C| \geq \frac{n}{2-\varepsilon}$, then $\Delta(G[C]) \geq \varepsilon d$. (This is known as a 'supersaturation' statement: if C is too big, then C must be 'far' from independent, which here is measured by $\Delta(G[C])$. Supersaturation is a necessary ingredient for a container lemma.) Show that the 2 cannot be replaced with 2.1.

---

[7] To check your work, the answer is $\sim 2^{2^d} \cdot 6 \cdot e$.

[8] To check your work, the answer is $\sim 6^{2^{d-1}} \cdot 20 \cdot \exp\left((4/3)^{d-1} + \frac{1}{3}\right)$.

[9] To check your work, the answer is $\sim 2^{2^{d-1}} \cdot 2 \cdot e$.

[10] To check your work, the answer is $\sim 4^{2^d} \cdot 4 \cdot \exp\left(\frac{1}{4}\left(\frac{3}{2}\right)^d + \frac{d(d+1)}{36}\left(\frac{9}{8}\right)^d + \frac{1}{4}\right)$.

[11] To check your work, the answer is $i(Q_d) = (1 + 2^{-d}\frac{3d^2-3d-2}{8} + o(2^{-d}))2\sqrt{e}2^{2^{d-1}}$.

(b) Let $\varepsilon > 0$. Greedily construct $S \subseteq V(G)$ such that $|S| \le \frac{n}{\varepsilon d}$ and $C = V(G) \setminus (S \cup N(S))$ satisfies $|C| \le \frac{n}{2-\varepsilon}$.[12]

(c) Let I be an independent set of G. Show that for every $S \subseteq I$, we have $I \setminus S \subseteq V(G) \setminus (S \cup N(S))$.

(d) Let $\varepsilon > 0$, and let I be an independent set of G. Greedily construct $S \subseteq I$ and C *which depends only on S, not I* such that $|S| \le \frac{n}{\varepsilon d}$, $I \setminus S \subseteq C$, and $|C| \le \frac{n}{2-\varepsilon}$. (Formally, construct functions $f : \mathcal{I}(G) \to \binom{|V(G)|}{\le n/\varepsilon d}$ and $g : \binom{|V(G)|}{\le n/\varepsilon d} \to \binom{|V(G)|}{\le n/(2-\varepsilon)}$ such that for every $I \in \mathcal{I}(G)$, $S \subseteq I$ and $I \setminus S \subseteq C$, where $S = f(I)$ and $C = g(f(I))$.)[13] (This algorithm is called the graph container algorithm; S is known as the 'certificate.')

(e) Finish the proof of Lemma 6 by running the algorithm from (d) on all independent sets of G.

4. ⋆⋆ Modify Lemma 6 and its proof (steps (a)-(f)) to produce containers for 'nearly' independent sets $I \subseteq V(G)$ with $\Delta(G[I]) \le b$.

5. ⋆⋆⋆ Refine the analysis of the container algorithm to obtain the following improved version of Lemma 6:[14][15]

Lemma 7. *Let G be an n-vertex d-regular graph. There exists $\mathcal{C} \subseteq 2^{V(G)}$ such that*

- *for every independent set I of G, there exists $C \in \mathcal{C}$ such that $I \subseteq C$,*
- $|\mathcal{C}| \le \binom{n}{\le \frac{n}{d}\log_2(d)}$, *and*
- *for every $C \in \mathcal{C}$, $|C| \le \frac{n}{d}\log_2(d) + \frac{n}{2}$.*

Use Lemma 7 to show that $\log i(G) \le \big(1 + O(\log^2(d)/d)\big)\, n/2$ for every n-vertex d-regular graph G. (Recall that Lemma 6 gives $O(\sqrt{\log(d)/d})$ as the error term, and the entropy argument gives the optimal $O(1/d)$ error.)

[12] Use (a).

[13] Taking a cue from (b), we start with $S = \emptyset$, $C = V(G) \setminus (S \cup N(S))$, iteratively find a highest degree vertex v of $G[C]$, and add v to S. But since we require $S \subseteq I$, we may only put v to S if $v \in I$. If $v \notin I$, then we do not add v to S, but we do delete v from C. While it appears that the C constructed here depends on I, you should argue that C only depends on S.

[14] Change the exit condition of the container algorithm from $\Delta(G[C]) \le \varepsilon d$ to $|C| \le n/2$.

[15] Iteratively analyze the graph container algorithm, considering several phases, each of which cuts $\Delta(G[C])$ by half.

Problem Set 1

1. $\star$ Recall this theorem from the lecture: if G is an n-vertex d-regular bipartite graph, then $\log_2 i(G) \leq (1 + O(1/d))\, n/2$. Give an example which shows that the dependence on d is best possible. ($\star\star$ If we've prove this theorem already, try to follow the proof on your example and notice how the inequalities become sharp.)

2. $\star - \star\star$ You might skim through this exercise if you are already familiar with Shannon entropy. In what follows $\mathbf{X}, \mathbf{Y}, \ldots$ are finitely-supported random variables. Recall that the (binary) entropy of $\mathbf{X}$ is

$$H(\mathbf{X}) = \sum_x \mathbb{P}(\mathbf{X} = x) \log \frac{1}{\mathbb{P}(\mathbf{X} = x)},$$

where the log is taken base 2. Entropy is a measure of the 'information' stored in a random variable, here measured in bits; another interpretation is that entropy is the average 'surprise' upon learning the realization of a random variable. Verify the following starting with only the above definition of entropy.

   (a) *(Uniform maximizes entropy)* Suppose $\mathbf{X}$ is supported on a finite set S. Show that $H(\mathbf{X}) \leq \log |S|$, with equality if and only if $\mathbf{X}$ is uniform on S. (This is why entropy is helpful for counting problems: if we want to count $|S|$, we may equivalently calculate the entropy of the uniform random variable on S.)[16]

   (b) *(Chain rule)* Recall the definition of conditional entropy:

$$H(\mathbf{X}|\mathbf{Y}) = \sum_y \mathbb{P}(\mathbf{Y} = y) \sum_x \mathbb{P}(\mathbf{X} = x|\mathbf{Y} = y) \log \frac{1}{P(\mathbf{X} = x|\mathbf{Y} = y)}.$$

   Show that

$$H(\mathbf{X}, \mathbf{Y}) - H(\mathbf{X}) = H(\mathbf{Y}|\mathbf{X}).$$

   (c) *(Additivity for independent variables)* We say that $\mathbf{X}$ and $\mathbf{Y}$ are independent if $\mathbb{P}(\mathbf{X} = x, \mathbf{Y} = y) = \mathbb{P}(\mathbf{X} = x)\mathbb{P}(\mathbf{Y} = y)$ for all x and y. Show that if $\mathbf{X}$ and $\mathbf{Y}$ are independent, then $H(\mathbf{Y}|\mathbf{X}) = H(\mathbf{Y})$ and $H(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}) + H(\mathbf{Y})$.

   (d) *(Dropping conditioning)* Show that $H(\mathbf{Y}|\mathbf{X}) \leq H(\mathbf{Y})$, and characterize equality.[17]

   (e) *(Subadditivity)* Show that $H(\mathbf{X}, \mathbf{Y}) \leq H(\mathbf{X}) + H(\mathbf{Y})$. Furthermore $H(\mathbf{X}) \leq \sum_i H(\mathbf{X}_i)$.

   (f) *(Dropping conditioning, part 2)* Show that $H(\mathbf{Y}|\mathbf{X}, \mathbf{Z}) \leq H(\mathbf{Y}|\mathbf{X})$.

   (g) *(Inserting conditioning)* Show that $H(\mathbf{X}|\mathbf{Z}) \leq H(\mathbf{X}|\mathbf{Y}) + H(\mathbf{Y}|\mathbf{Z})$.

[16] There is an elementary proof using the inequality $\log(x) \leq x - 1$. Alternatively, one can use Jensen's inequality.

[17] Use $\log(x) \leq x - 1$ or Jensen's Inequality *judiciously*, in addition to Bayes' Rule.

(h) *(Refinement)* If $\mathbf{Y}$ determines $\mathbf{X}$, then $H(\mathbf{X}|\mathbf{Y}) = 0$ and $H(\mathbf{Z}|\mathbf{Y}) \le H(\mathbf{Z}|\mathbf{X})$.

(i) *(Shearer's Inequality)* Let $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_k)$ be a random vector, and let $\alpha : 2^{[k]} \to \mathbb{R}_{\ge 0}$ be such that $\sum_{A \ni i} \alpha(A) \ge 1$ for all $i \in [k]$. Show that

$$H(\mathbf{X}) \le \sum_{A \subseteq [k]} \alpha(A) H(\mathbf{X}_A),$$

where $\mathbf{X}_A = (\mathbf{X}_i)_{i \in A}$.[18]

3. ⋆⋆ Let $0 \le \alpha \le 1/2$. Show that the number of subsets of $[n]$ of size at most αn is at most $2^{h(\alpha)n}$, where $h(\alpha) = \alpha \log_2\left(\frac{1}{\alpha}\right) + (1-\alpha)\log_2\left(\frac{1}{1-\alpha}\right)$.[19] Derive that $\binom{n}{\le k} \le \left(\frac{en}{k}\right)^k$.[20]

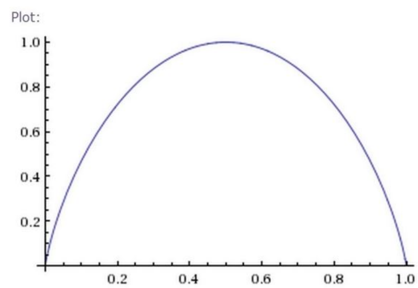

Figure 1: a plot of $h(\alpha)$

4. ⋆⋆ Find a bijection between the following two collections.

- (proper) 3-colorings of Q_d with the color of a prefixed vertex v_0 is fixed;
- the graph homomorphisms from Q_d to $\mathbb{Z}$ (in which two vertices a, b are adjacent iff $|a - b| = 1$) with $f(v_0) = 0$.

[18] For each $A \subseteq [k]$, we have $H(\mathbf{X}_A) = \sum_{i \in A} H(\mathbf{X}_i | \mathbf{X}_{A \cap [i-1]}) \ge \sum_{i \in A} H(X_i | X_{[i-1]})$.

[19] Image from D. Galvin, *Three tutorial lectures on entropy and counting*, arXiv 1406.7872

[20] Where have you seen $h(\alpha)$ before? Express the counting problem as an entropy problem and use subadditivity.