Problem Set 8 CS265, Fall 2022

Due: December 9 (Friday) at 23:59 (Pacific Time)

Group members: Rylan Schaeffer, Connor Settle

Please follow the homework policies on the course website.

---

1. **(5 pt.) Sampling *Without* Replacement**

   Suppose there are n total balls, of which m are marked. We sample k many of the balls uniformly *without replacement*.[1] Let Z be the random variable denoting how many of the k balls are marked. In this problem, you will show that Z is concentrated around its mean.

   (a) **(1 pt.)** Show that $\mathbb{E}[Z] = \frac{km}{n}$.

   (b) **(4 pt.)** When $k \geq 1$, show that $\Pr[|Z - \mathbb{E}[Z]| \geq \lambda] \leq 2e^{-\lambda^2/(2k)}$ for any $\lambda > 0$.

   [**HINT:** *Try applying the Azuma-Hoeffding tail bound to a Doob martingale. When applying Azuma-Hoeffding to a martingale $\{Z_t\}$, feel free to provide a short/intuitive explanation for why $|Z_i - Z_{i-1}| \leq c_i$ rather than a rigorous proof.*]

   (c) [**Optional: this won't be graded.**] When k is close to n, a tighter bound than that from part (b) holds.

   i. **(0 pt.)** When $k = n$, explain why $\Pr[Z = \mathbb{E}[Z]] = 1$.

   ii. **(0 pt.)** When $1 \leq k \leq n-1$, show that $\Pr[|Z - \mathbb{E}[Z]| \geq \lambda] \leq 2e^{-\lambda^2/(2v)}$ where v is defined as

   $$v := \sum_{i=1}^{k} \left(1 - \frac{k-i}{n-i}\right)^2.$$

   iii. **(0 pt.)** Show that $v \leq O(k(n-k)/n)$. This shows that the bound from part (c), ii is tighter than the bound from part (b) when k is close to n.

SOLUTION:

(a) Let X_i denote that the ith ball was marked. Then:

$$\mathbb{E}[Z] = \mathbb{E}[\sum_{i=1}^{k} X_i] = k\mathbb{E}[X_i] = k\frac{m}{n}$$

(b) Let A be the number of marked balls out of the k balls we will draw. Our Doob martingale will then be $Z_t = \mathbb{E}[A|X_{\leq t}]$. Note that each $|Z_i - Z_{i-1}| \leq 1$ because either Z_{i-1} increments (if we draw another marked ball on the ith index) or doesn't (if we don't draw another marked ball on the ith index). Then, per Azuma-Hoeffding,

$$\mathbb{P}[|Z - \mathbb{E}[X]| \geq \lambda] \leq 2\exp\Big(-\frac{\lambda^2}{2\sum_{i=1}^{k} c_i}\Big) = \exp\Big(-\frac{\lambda^2}{2k}\Big)$$

[1] Note that this only makes sense when $k, m \leq n$.

2. **(11 pt.) Homework Solution Consensus**

Suppose that your homework group[2] is working on a very difficult multiple choice question, where only one of the answers is correct. It turns out that your opinion on which choice is correct differs from the opinions of many other members of the group. Fortunately, you have many friends in this group who are willing to listen to your opinion, and you are willing to listen to theirs as well. You want to talk with your friends hoping that all the group members will eventually agree on the same choice.

Formally, there is an undirected graph $G = (V, E)$ whose vertices represent the group members and a pair of members are friends if and only if they are connected by an edge. For simplicity, we assume that G contains none of the following: 1) self-loops, 2) multiple edges connecting the same pair of vertices, or 3) isolated vertices, i.e., vertices with no edge on them. Let S be the set of possible answers to the homework question (for example, $S = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\}$). We can represent the opinions of the group members by a mapping $\sigma : V \to S$ where the group member corresponding to vertex v thinks that $\sigma(v)$ is the correct answer.

The opinions σ of the group members evolve due to discussions between friends. We model the evolution of σ by the following time-homogeneous Markov chain: starting from the initial opinion σ_0, σ changes from σ_{t-1} to σ_t at step t as follows. Independently for every vertex v, we flip a fair coin. If the outcome is "heads", $\sigma_t(v)$ remains the same as $\sigma_{t-1}(v)$; otherwise, $\sigma_t(v)$ becomes $\sigma_{t-1}(v')$ for a uniformly random neighbor v' of v. In short, every group member keeps their own opinion with probability 1/2, and takes one of their friends' opinion with the remaining 1/2 probability.

In this problem, we will determine the likelihood that the group members reach a certain consensus, given their initial opinions.

(a) **(1 pt.)** If G is disconnected and $|S| > 1$, show that there exist initial opinions σ_0 of the members for which consensus is never reached.

(b) **(3 pt.)** If G is connected, show that consensus is eventually reached almost surely. That is, show that as the number of steps goes to infinity, the probability that consensus has been reached approaches 1.

(c) **(2 pt.)** Let X_t be the number of group members who think that choice A is the correct answer after step t. Give an example where $(X_t)_{t\geq 0}$ is *not* a martingale with respect to $(\sigma_t)_{t\geq 0}$. The example should be one specific tuple (G, S, σ_0).

(d) **(3 pt.)** Let Y_t be the sum of the degrees of the vertices v corresponding to the group members who think that choice A is the correct answer after step t. Prove that $(Y_t)_{t\geq 0}$ is a martingale with respect to $(\sigma_t)_{t\geq 0}$.

(e) **(2 pt.)** Assume that G is connected. What is the probability that every member of the group eventually thinks that choice A is the correct answer? Express your answer in terms of G and the initial opinion σ_0 of the group members.

[**HINT:** *Try applying the martingale stopping theorem to the martingale* $(Y_t)_{t\geq 0}$.]

[2]For the purposes of making this question more interesting, pretend that your homework group has more than three people in it...

SOLUTION:

(a) Consider G with two connected subgraphs G_1 and G_2 with all vertices in G_1 initialized such that $\forall v_1 \in G_1, \sigma_0(v_1) = s_1 \in S$ and with all vertices in G_2 initialized such that $\forall v_2 \in G_2, \sigma_0(v_2) = s_2 \in S$, with the key constraint that $s_1 \neq s_2$. In this graph, consensus will never be reached because each subgraph can never "cross-pollinate" their respective solutions to the other subgraph.

(b) First, note that if consensus is reached, consensus can never be escaped because if anyone tries to change their opinion, all surrounding opinions match their opinion, so randomly sampling a replacement opinion will return the original opinion.

Next, consider some connected G at some point in time t. If consensus has been reached, the probability of consensus is 1. If consensus hasn't been reached, we have to do a bit more work. In the worst case, G could be a "line" (I don't know the technical term) such that information takes $|V| - 1$ steps to transmit from one side to the other side. In the worst case, consider the left vertex trying to transmit its solution to all nodes to its right. There is a small, but non-zero probability that (1) the consensus grows left to right, and that (2) along the way, the consensus group doesn't change their mind. Let E be the event that such propagation occurs to reach consensus, and let is occur with probability $p > 0$.

This E is a geometric random variable with $p > 0$, whose CDF approaches 1 almost surely. Thus consensus is eventually reached almost surely.

(c) Define G as a fully connected graph with 3 nodes, $S = \{A, B\}$, and σ_0 with v_1 holding opinion A and v_2, v_3 holding opinions B. Note that $X_0 = 1$ since 1 group member holds opinion A. In order to qualify as a martingale, $\mathbb{E}[X_1|\sigma_0]$ must equal $X_0 = 1$. What is the expected value of X_1? There are 8 possible outcomes of σ_1 for all 2^3 possible vertex-opinion pairings. If you work out the probabilities, the expected value of $\mathbb{E}[X_1] < 1$, intuitively because we expect B to dominate A since more people hold opinion B. In general, any "imbalanced" graph will make $(X_t)_{t \geq 0}$ not a martingale with respect to $(\sigma_t)_{t \geq 0}$ because we expect X_t to either grow (if A is the majority opinion) or shrink (if A is the minority opinion).

(d) First, note that Y_t is a function of $(\sigma_{t'})_{0 \leq t' \leq t}$ because Y_t specifically requires knowledge of σ_t to be computed. Second, note that $Y_t < \infty$ because Y_t is at most $|V| *$ max degree of $G < |V|^2 < \infty$. Third, we want to show that

$$\mathbb{E}[Y_t|(\sigma_{t'})_{0 \leq t' < t}] = Y_{t-1}$$

Let d^v denote the degree of vertex $v \in V$. Starting with the left-hand size, the expected value of Y_t is the degree of each vertex times the probability the vertex takes opinion A:

$$\mathbb{E}[Y_t|(\sigma_{t'})_{0 \leq t' < t}] = \sum_v d^v \, \mathbb{P}[\sigma_t(v) = A|\sigma_{t-1}]$$

We can break the sum up into two components: at time step $t - 1$, either v had opinion A or v did not have opinion A:

$$\mathbb{E}[Y_t|(\sigma_{t'})_{0 \leq t' < t}] = \sum_{v:\sigma_{t-1}(v)=A} d^v \, \mathbb{P}[\sigma_t(v) = A|\sigma_{t-1}] + \sum_{v:\sigma_{t-1}(v)\neq A} d^v \, \mathbb{P}[\sigma_t(v) = A|\sigma_{t-1}]$$

Let n^v_{t-1} denote the number of neighbors of v at time $t-1$ with opinion A i.e. $\sigma_{t-1}(v) = A$. In the case that v had opinion A, the probability v keeps opinion A is the sum of the probability that v doesn't change i.e. $(1/2)$, plus the probability that v changes but samples a neighbor with opinion A i.e. $(1/2) * (n^v_{t-1}/d^v)$. In the case that v did not have opinion A, the probability that it changes to opinion A is $(1/2) * (n^v_{t-1}/d^v)$. Together:

$$\mathbb{E}[Y_t|(\sigma_{t'})_{0\leq t'<t}] = \sum_{v:\sigma_{t-1}(v)=A} d^v \, (\frac{1}{2} + \frac{n^v_{t-1}}{2d^v}) + \sum_{v:\sigma_{t-1}(v)\neq A} d^v \, (\frac{n^v_{t-1}}{2d^v})$$

Rearranging, we have:

$$\mathbb{E}[Y_t|(\sigma_{t'})_{0\leq t'<t}] = \sum_{v} \frac{n^v_{t-1}}{2} + \sum_{v:\sigma_{t-1}(v)=A} \frac{d^v}{2}$$

The right term is $Y_{t-1}/2$ by definition of Y_{t-1}, and the left term is also equal to $Y_{t-1}/2$ because n^v_{t-1} is 0 if v doesn't have opinion A. The sum is therefore Y_{t-1} and we have proved the goal:

$$\mathbb{E}[Y_t|(\sigma_{t'})_{0\leq t'<t}] = Y_{t-1}$$

(e) Note that we can apply the Martingale Stopping Theorem to $(Y_t)_t$ under condition 1, because we previously showed that $Y_t < |V|^2$. The theorem then tells us that for random non-negative integer $T \geq 0$:

$$\mathbb{E}[Y_T] = \mathbb{E}[Y_0] = Y_0$$

where the last equality follows because $Y_0 = Y_0(G, \sigma_0)$ is a deterministic quantity computable from G and σ_0. Second, define $M =$ the sum over all vertices of each vertex's degree. Note that $Y_t \in [0, M]$ since at least no one has opinion A or at most everyone has opinion A. Let T denote the first time Y_T hits either M or 0. Then from the stopping theorem and the definition of the expected value:

$$\mathbb{E}[Y_0] = \mathbb{E}[Y_T] = 0 * \mathbb{P}[Y_T = 0] + M\mathbb{P}[Y_T = M]$$

Rearranging, we have that:

$$\mathbb{P}[Y_T = M] = \frac{\mathbb{E}[Y_0]}{M}$$

This intuitively tells us that the probability opinion A dominates is proportional to the fraction in the starting graph.