

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

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 .

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

Suppose that your homework group² 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, $S = \{A, B, C, D\}$). We can represent the opinions of the group members by a mapping $\sigma : V \rightarrow S$ where the group member corresponding to vertex v thinks that $\sigma(v)$ is the correct answer.

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

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

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 $(Y_t)_{t \geq 0}$.]