# Final Examination (Take Home)

Out: Thursday, April 10, 2025
Due: Tuesday, April 24, 2025

**Exam instructions (please read carefully):**

- **Allowed resources.** You are allowed to consult the course web site, your notes, and the course textbook. You are not allowed to use other resources, such as other books, other web sites, or other people.
- **Non-collaboration.** No collaboration is permitted on this exam. It is trusted that you will not discuss this exam or related course material with any other person (classmate or otherwise). You must abide by the Brown University Academic Code concerning examinations, quizzes, and tests (see `http://goo.gl/mQtfSa`).
- **Questions.** There will be no TA hours April 10-24. For any clarification of possible textual ambiguities email `cs1550tas@lists.brown.edu` with your question. *We will post clarifications (anonymously) to Ed Discussion.*
- **Handing in.** Submit to Gradescope by the due date.

**PLEASE SIGN (failure to sign voids the exam)**: I solemnly state that I have abided by the exam instructions stated above, including the tenets of the Brown University Academic Code concerning examinations, quizzes and tests.

**Signature:** ______________________________

**Name (please print):** ______________________________

### GENERAL EXAM INSTRUCTIONS ON PREVIOUS PAGE

## Problem 1 (20 points)

Suppose we are given an $n$-vertex undirected graph $G = (V, E)$, where the vertices are numbered $1, 2, \ldots, n$. An independent set of $G$ is a subset $W \subseteq V$ of the vertices such that there is not an edge $e \in E$ connecting two vertices in $W$. We wish to find a (hopefully large) independent set in $G$. Given a permutation $\sigma$ of the vertices, define a subset $S(\sigma)$ of the vertices as follows: for each vertex $i$, $i \in S(\sigma)$ if and only if no neighbor $j$ of $i$ precedes $i$ in the permutation $\sigma$.

1. Show that each $S(\sigma)$ is an independent set in $G$.
2. Suggest a natural randomized algorithm to produce $\sigma$ for which you can show that the expected cardinality of $S(\sigma)$ is
$$\sum_{i=1}^{n} \frac{1}{d_i + 1},$$
where $d_i$ denotes the degree of vertex $i$.
3. Prove that $G$ has an independent set of size at least $\sum_{i=1}^{n} \frac{1}{d_i+1}$.

## Problem 2 (20 points)

Consider the following model for a digital communication channel with noise. Alice wants to communicate a message $m$ to Bob. First, Alice encodes $m$ as a string $s$ of length $n$, using an error-correcting code. This means that it is enough for Bob to have $n(1-q)$ correct bits of $s$ (where $q$ is a constant between 0 and 1) in order to correctly decode it and obtain Alice's original message $m$. That is, if Bob has a string $s'$ that matches $s$ in at least $n(1-q)$ places, then Bob can recover $m$.

Each bit that Alice sends over the channel is received correctly by Bob with probability $p > 1/2$, independently of other bits. Alice wants to send $s$ over the channel, but to improve the probability that Bob can recover $m$, Alice sends $s$ several times and Bob takes the majority value for each bit to obtain $s'$. Bob then attempts to recover $m$ from $s'$.

1. Fix a constant $r \in (0, 1)$. How many times does Alice need to send $s$ so that the probability that a given bit of $s$ has a correct majority is at least $1 - r$?
2. Now, for which values of $r \in (0, 1)$ does Alice need to use so that Bob can recover $m$ correctly with probability at least $1 - \epsilon$? You can give an inequality involving $r$, instead of solving for $r$ explicitly.

## Problem 3 (20 points)

1. Given two range spaces $(X, \mathcal{R}_1)$ and $(X, \mathcal{R}_2)$, each with VC-dimension $d$, we define a "union" range space $(X, \mathcal{R}_{1\cup 2})$:
$$\mathcal{R}_{1\cup 2} = \{r_1 \cup r_2 \mid r_1 \in \mathcal{R}_1 \text{ and } r_2 \in \mathcal{R}_2\}.$$
Prove that the VC dimension of $\mathcal{R}_{1\cup 2}$ is $O(d)$.

   You may use without proof the result of Exercise 14.9: for $n \geq 2d$ and $d \geq 1$, the growth function satisfies $G(d, n) = \sum_{i=0}^{d} \binom{n}{i} \leq 2 \left(\frac{ne}{d}\right)^d$. (You may not use Theorem 14.5 or Corollary 14.6 directly, because those results are not proven in the book.)

2. The Martian citizens are electing a new leader and Jupiter's evil team is trying to influence the election results. To maximize its influence the Jupiter team wants to identify the 5 circles of Martian population most likely to vote for candidate Good.

   Let $C = \{S_1, \ldots, S_5\}$ be a set of 5 circles (each $S_i$ is a circle). For a location (point) $X$ on the map, we denote $X \in C$ if $X \in \cup_{i=1}^{5} S_i$. We also denote $P$ as the set of all Martian locations planning to vote for candidate Good. Define the **error** of $C$ as

$$\mathrm{err}(C) = \Pr(X \in P, X \notin C) + \Pr(X \notin P, X \in C),$$

   which measures how well $C$ minimizes the coverage of places not voting for Good and maximizes the coverage of locations voting for Good. Let $C^*$ be the set of the 5 circles with minimum error. The Jupiter team can sample uniformly at random a locations in the country and check who they plan to vote for. For a given $\epsilon > 0$, their goal is to find a set $C'$ such that $\mathrm{err}(C') \leq \mathrm{err}(C^*) + \epsilon$. How many samples do they need in order to find such a set of 5 circles with probability at least $1 - \delta$, for $\delta \in (0, 1)$? You can give the sample complexity using big $O(\cdot)$ notation. [You can use the known fact that all Martians live in a 2-D planar region.]

# Problem 4 (20 points)

Consider a random walk $Z_t = (X_t, Y_t)$ on the the 2-dimension integer grid. The walk starts at point $Z_0 = (0, 0)$, and at each step the walk moves with equal probabilities one step left, right, up or down. I.e.

$$Pr(Z_t = (x, y) \mid Z_{t-1} = (x', y')) = \begin{cases} 1/4 & \text{if } x = x' + 1 \text{ and } y = y' \\ 1/4 & \text{if } x = x' - 1 \text{ and } y = y' \\ 1/4 & \text{if } x = x' \text{ and } y = y' + 1 \\ 1/4 & \text{if } x = x' \text{ and } y = y' - 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Answer the following questions.
   (a) Show that each of the sequences $X_0, X_1, \ldots$ and $Y_0, Y_1, \ldots$ are martingales.
   (b) Let $A_t = X_t + Y_t$ for $t \geq 0$. Show that the sequence $A_0, A_1, \ldots$ is a martingale.
2. Fix a value $t \geq 0$. Define a (small) set $S(t)$ of points on the grid, such that

$$Pr(Z_t \notin S) \leq \delta < 1.$$

   The total number of points in $S(t)$ should be $O(t \ln(1/\delta))$.

# Problem 5 (20 points)

1. We prove an extension of the Chernoff bound that was implicitly assumed in class. Let $X = \sum_{i=1}^{n} X_i$, where the $X_i$'s are independent 0-1 random variables. Let $\mu = E[X]$. Choose any $\mu_L$ and $\mu_H$ such that $\mu_L \leq \mu \leq \mu_H$. Prove the following:
   (a) For any $\delta > 0$,

$$Pr(X \geq (1 + \delta)\mu_H) \leq \left( \frac{e^{\delta}}{(1 + \delta)^{(1+\delta)}} \right)^{\mu_H}.$$

   (b) For any $0 < \delta < 1$,

$$Pr(X \leq (1 - \delta)\mu_L) \leq \left( \frac{e^{-\delta}}{(1 - \delta)^{(1-\delta)}} \right)^{\mu_L}.$$

[Hint: Contract random variable $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ such that for all $1 \leq i \leq n$, we always have $Y_i \leq X_i \leq Z_i$, and their sums have the required expectations.]

2. An $(\epsilon, \delta)$-confidence interval for an unknown constant $T$ is a (random) interval $[A - \epsilon, A + \epsilon]$ such that

$$Pr(T \notin [A - \epsilon, A + \epsilon]) \leq \delta.$$

Let $X_1, \ldots, X_n$ be $n$ independent observations such that $X_i = 1$ with unknown probability $P$, otherwise $X_i = 0$. Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Use the first part of this question to bound the error probability $\delta$ of the confidence interval $[\bar{X} - \epsilon, \bar{X} + \epsilon]$ for $P$.