

Final Review

CS70: Discrete Mathematics and Probability Theory

UC Berkeley – Summer 2026

Final Exam – Student Request Topics

- Graph/edge coloring

Vertex and edge coloring problems are good sources of problems to check your intuition of basic graph properties. Let's look at some problems on the board, including:

- How is vertex coloring related to degree? Can you think of a graph with max degree 2 that requires 3 colors? Can you think of a graph with max degree 100 that only needs 2 colors? How many colors does K_n require?
- Combining graphs: Let G_1 and G_2 be two graphs that can each be colored with 2 colors. If you connect them with a single edge to create a new graph G , how many colors do you need for G ? What if you connect with *two* edges?
- Edge coloring: For examples above, how many edge colors are needed?

Final Exam – Student Request Topics Continued

- Order statistics / the CDF method (e.g. Fa24 14.1–14.4, Fa23 15): going from max/min of independent uniforms to a CDF, then differentiating to a PDF.

Let's work through 14.1 and 14.2 together (14.3 and 14.4 follow easily): X, Y are independent and uniform on $[0, r]$, and let $Z = \max(X, Y)$ – what are the CDF and PDF of Z ?

- First-step equations, specifically which system to write (e.g. Fa24 20.1 and 21.2, Fa23 17): when it's $\beta = 1 + P\beta$ vs $\alpha = P\alpha$ with no +1 vs $\pi = \pi P$, and where the +1 comes from.

This will be in the following review slides...

- Counting solutions to $ax \equiv b \pmod{pq}$ (e.g. Fa24 6.4): counting invertible a , and how the solution count changes when $\gcd(a, pq)$ isn't 1.

Let's work this out when $\gcd(a, pq) = 1$ and when $\gcd(a, pq) = p$.

- The abstract Δ_i polynomials and Berlekamp-Welch (e.g. Fa24 7.3).

These are Lagrange polynomials for interpolation. Let's look at $\Delta_i(i)$ and $\Delta_i(j)$ (for j in set of x values).

Lecture 13 Summary – Counting II

Combinatorial Proofs: Identity from counting same set in two ways

Basic combinations: $\binom{n}{k} = \binom{n}{n-k}$ ways to include k = ways to exclude $n - k$

Binomial Coefficients and Binomial Theorem

Expansion of monomial power is ways of choosing a factors

Pascal's Triangle and Pascal's Identity: $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$.

LHS: Count as number of subsets of $n + 1$ items size k .

RHS: $\binom{n}{k-1}$ counts subsets of $n + 1$ items with first item.

$\binom{n}{k}$ counts subsets of $n + 1$ items without first item.

Disjoint – so add!

Hockey Stick Identity

Count as subsets, and count as subsets with given smallest element

Inclusion/Exclusion: Two sets of objects

Add number of each subtract intersection of sets

Sum Rule: If intersection is empty, nothing to subtract!

Inclusion/Exclusion: General

Alternate adding/subtracting: Add 1-way, subtract 2-way, add 3-way, ...

Concept Check

What “two ways of counting” give a combinatorial proof that $k\binom{n}{k} = n\binom{n-1}{k-1}$?

Counting 1: Choose team of k players from n , then pick captain from the k members.

Counting 2: Choose a captain from n choices, then the other $k - 1$ team members.

For three sets A , B , and C , how many intersection subsets do you subtract and add from $|A| + |B| + |C|$ to get $|A \cup B \cup C|$?

3 subtractions ($|A \cap B|$, $|A \cap C|$, and $|B \cap C|$); one addition ($|A \cap B \cap C|$).

Lecture 14 Summary – Countability

Sizes of sets

- Comparing sizes of two sets – bijections
- Infinite sets too!

Countably infinite sets

- Counting numbers ($\mathbb{N}$) are countable (surprise!)
- Integers are countable
- Rationals are countable
- The set of all finite-length binary strings is countable

Uncountable sets

- The set of reals is uncountable
- Diagonalization as a proof technique
- The set of subsets of $\mathbb{N}$ is uncountable

Bottom line: Infinity is weird. And cool.

Concept Check

Countable or uncountable?

- The set of all finite length strings of English letters.

Countable.

- The set of all subsets of integers.

Uncountable.

- The set of all finite length decimal representations of real numbers.

Countable. Note that this is *not* the set of all real numbers! It excludes some obvious numbers, like π and $\sqrt{2}$, but also $\frac{1}{3}$.

True or false: Every surjective (onto) mapping from integers to even integers is a bijection.

False. There *is* a bijection, but not every mapping is a bijection. For example, $f(x) = 2\lfloor x/2 \rfloor$ is surjective, but is not a bijection.

Lecture 15 Summary – Computability

The problem with self-reference

Liar's Paradox, Russell's Paradox, ...

Uncomputable functions

Programs as data

The halting problem

Proof of undecidability

Diagonalization view of the proof

Some History

Alan Turing

David Hilbert

Reductions: Solving one problem with another

Undecidability of basic program behavior

HALT reduction to “prints hello world”

Some additional uncomputable functions

Gödel's Incompleteness Theorem

Concept Check

- What basic proof technique is at the heart of the undecidability proof of the halting problem?

Proof by contradiction. You might say “diagonalization” but that depends on whether that is a “basic proof technique” (I’d say not...).

- True or false: We can write a program analyzer that takes a program as input and always correctly determines if the program performs a division by zero when it is run.

False. Basically any analyzer of non-trivial program *behavior* is not possible.

- True or false: We can write a program analyzer that takes a program as input and always correctly determines if the program follows style guidelines for variable names.

True. This just requires scanning the program for variable names and checking if they follow guidelines. This does not depend on the runtime behavior of the program at all.

Lecture 16: Summary – Discrete Probability Basics

1 Random Experiment

2 Probability Space: Ω ; $\mathbb{P}[\omega] \in [0, 1]$; $\sum_{\omega} \mathbb{P}[\omega] = 1$

3 Uniform Probability Space: $\mathbb{P}[\omega] = 1/|\Omega|$, for all $\omega \in \Omega$

4 Event, “subset of outcomes”: $A \subseteq \Omega$, and $\mathbb{P}[A] = \sum_{\omega \in A} \mathbb{P}[\omega]$

$\Rightarrow$ On a uniform space (common!), just counting $|A|$ since $\mathbb{P}[A] = \frac{|A|}{|\Omega|}$

5 Examples!

Concept Check

True or false:

- For events A and B , it's possible that $\mathbb{P}[A] + \mathbb{P}[B] > 1$.

True. Events are just subsets and don't have to be proper subsets, so we could in fact have $A = B = \Omega$, and then $\mathbb{P}[A] + \mathbb{P}[B] = 2$.

- The sample space Ω always has finite size.

False. Think of the “flips until heads” example.

For the experiment of rolling two six-sided dice, how many points are in the sample space?

36. Every outcome is of the form (a, b) where a and b are in the range 1 to 6.

Lecture 17 Summary – Combinations of Events

Topics:

- Conditional Probability and Bayesian Inference

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

- Total Probability

$$\mathbb{P}[A] = \mathbb{P}[A|B] \cdot \mathbb{P}[B] + \mathbb{P}[A|\bar{B}] \cdot \mathbb{P}[\bar{B}]$$

- Bayes' Rule

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[B|A] \cdot \mathbb{P}[A]}{\mathbb{P}[B]}$$

- Independent Events and Event Intersection

$$\text{Independent } A \text{ and } B \iff \mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$$

- Intersection for Non-Independent Events (Product Rule)

$$\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A] \cdot \mathbb{P}[B|A] \cdot \mathbb{P}[C|A \cap B]$$

- Event Union and Inclusion-Exclusion

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$$

Note: Formulas are only simplest forms, as reminders

Concept Check

True or false?

- If events $B_1, \dots, B_k$ partition the sample space, then $\mathbb{P}[A \cap B_1] + \mathbb{P}[A \cap B_2] + \dots + \mathbb{P}[A \cap B_k] = 1$.

False. This adds up to $\mathbb{P}[A]$.

- If events $B_1, \dots, B_k$ partition the sample space, then $\mathbb{P}[B_1|A] + \mathbb{P}[B_2|A] + \dots + \mathbb{P}[B_k|A] = 1$.

True.

- For events A and B , $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$.

False. This is only true if the events are independent.

Under what condition(s) is $\mathbb{P}[A|B] = \mathbb{P}[B|A]$?

$\mathbb{P}[A] = \mathbb{P}[B]$, since Bayes theorem says $\mathbb{P}[A|B] = \mathbb{P}[B|A] \frac{\mathbb{P}[A]}{\mathbb{P}[B]}$.

$\mathbb{P}[A]$ and $\mathbb{P}[B]$ must also be non-zero for conditional probabilities to be defined.

Lecture 18 Summary – Applications

Considered three real applications – all really just “balls and bins!”

- Analyzing hash functions

For: data structures, file identification/integrity, cryptography, blockchain, ...

$$\mathbb{P}[\text{collision}] = p \approx e^{-m^2/2n}$$

Given any two of p , m , and n , can solve for the other

- Data coverage by random sampling

Exploring randomly – how long to see everything?

Coupon collector: Get all with probability $\geq 1/2$ with $n \ln(2n)$ samples

- Load balancing

Spreading out computational work or resources evenly

When $m = n$, probability is $\leq 1/2$ that worst-case bin has $> \frac{\ln n}{\ln \ln n}$ balls

Concept Check

For large n , if we draw m samples with replacement, approximately what should m be for there to be a 50-50 chance of a collision?

$m \approx \sqrt{n}$ is good enough. A little more precisely, $m \approx (\sqrt{2\ln 2})\sqrt{n}$.

True or False: Every child would love to collect cards featuring Turing Award winners.

Obviously true.

Lecture 19 Summary – Random Variables

- A random variable X is a function $X : \Omega \rightarrow \mathbb{R}$
- Preimages $X^{-1}(a)$ are just events
$$\mathbb{P}[X = a] := \mathbb{P}[X^{-1}(a)] = \mathbb{P}[\{\omega \mid X(\omega) = a\}]$$
- Distribution of X is the set of pairs: $\{(a, \mathbb{P}[X = a]), a \in \mathcal{A}\}$.
- Expected value: $\mathbb{E}[X] := \sum_a a \cdot \mathbb{P}[X = a]$
- Common distributions:
 - Bernoulli (success in one trial): $X \sim \text{Ber}(p)$
 - Binomial (success in n trials): $X \sim \text{Bin}(n, p)$
- Multiple random variables:
 - Joint distributions and marginal distributions
 - Independence
- Linearity of expectation: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- Examples: Rolling dice, flipping coins, fixed points of permutation, ...

Concept Check

True or False:

- A random variable can have values that are pairs of integers.
False. The value of a random variable is always a single real number.
- Events $X = 1$ and $X = 2$ with non-zero probabilities are independent.
False. $\mathbb{P}[X = 1 \cap X = 2] = 0$ while probabilities of individual events are non-zero.

The Binomial distribution is the sum of independent random variables with what distribution?

Bernoulli.

If $X \sim \text{Binomial}(10, p)$ and $Y \sim \text{Binomial}(10, q)$, and $Z = X + Y$, are X and Z independent? What is $\mathbb{E}[X + Z]$?

Not independent. Consider: If $X = 3$ then there's no way for $Z = 2$.
 $\mathbb{E}[X + Z] = \mathbb{E}[X + (X + Y)] = 2\mathbb{E}[X] + \mathbb{E}[Y] = 20p + 10q.$

Lecture 20 Summary

Conditional expectation

Basic properties: $\mathbb{E}[X | Y = y] = \sum_x x \cdot \mathbb{P}[X = x | Y = y]$

Law of Total Expectation: $\mathbb{E}[X] = \sum_y \mathbb{E}[X | Y = y] \cdot \mathbb{P}[Y = y]$

Functions of a random variable

Sums and products

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \quad ; \quad \text{when independent, } \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Variance

Definitions and meaning

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = (\mathbb{E}[X])^2 - \mathbb{E}[X^2]$$

Variance of common distributions: Bernoulli, Binomial, ...

Properties of variance

$$\text{Var}(X + c) = \text{Var}(X) ; \text{Var}(c \cdot X) = c^2 \cdot \text{Var}(X) ; \text{Ind: } \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Covariance

Definition and meaning: $\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Positively and negatively correlated random variables

Correlation: Scaled covariance: $\text{Corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$

Concept Check

True or False: $\text{Var}(cX) = c\text{Var}(X)$.

False (unless $c = 1$). In general, $\text{Var}(cX) = c^2\text{Var}(X)$.

For random variables X and Y , what condition guarantees that $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$?

X and Y should be independent.

For random variables X and Y , what is the maximum possible value of $\text{cov}(X, Y)$?

There is no finite maximum!

For random variables X and Y , what is the maximum possible value of $\text{Corr}(X, Y)$?

1.

Lecture 21 Summary

Added two more common distributions to our toolbox!

Geometric distribution: $X \sim \text{Geometric}(p)$

- $\mathbb{P}[X = k] = (1 - p)^{k-1}p$ (for $i \geq 1$) ; $\mathbb{E}[X] = \frac{1}{p}$; $\text{Var}(X) = \frac{1-p}{p^2}$
- Memoryless distribution
- Used to analyze expected number of steps in coupon collecting

Poisson distribution: $X \sim \text{Poisson}(\lambda)$

- $\mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}$ (for $i \geq 0$) ; $\mathbb{E}[X] = \lambda$; $\text{Var}(X) = \lambda$
- Events happen at a given expected rate
- Limiting distribution of increasingly fine-grained Binomial
- Sum of two Poisson is Poisson

Concept Check

Which distribution is memoryless?

Geometric.

Which distribution is useful to find the expected value of the coupon collector's problem?

Geometric. Collect until success (get the next distinct coupon/card).

The sum of two random variables with Poisson distributions with different parameters has what distribution?

Poisson.

The sum of two random variables with binomial distributions with different parameters has what distribution?

Not binomial...

The sum of two random variables with binomial distributions with the same parameters has what distribution?

Binomial (just more Bernoulli trials with same probability of success).

Lecture 22 Summary

Markov's Inequality: $\mathbb{P}[X \geq c] \leq \frac{\mathbb{E}[X]}{c}$

- X must be a non-negative random variable (i.e., $\mathbb{P}[X < 0] = 0$)
- All you need to know about X is $\mathbb{E}[X]$
- Best bound possible if all you know is $\mathbb{E}[X]$, but often not good enough

Chebyshev's Inequality: $\mathbb{P}[|X - \mu| \geq c] \leq \frac{\text{Var}(X)}{c^2}$

- Uses $\text{Var}(X)$ as indication of “spread” of distribution
- Best possible bound if all you know is $\mathbb{E}[X]$ and $\text{Var}(X)$
- Gives good results for estimating probabilities and polling
- Confidence intervals – estimating $\mathbb{E}[X]$ with average of n samples A_n :

$n = \frac{\text{Var}(X)}{\delta \varepsilon^2}$ guarantees $\mathbb{P}[|A_n - \mu| \geq \varepsilon] \leq \delta$

Estimating probability p (polling): Use $n = \frac{1}{4\delta \varepsilon^2}$

(Weak) Law of Large Numbers $\lim_{n \rightarrow \infty} \mathbb{P}[|A_n - \mu| \geq \varepsilon] = 0$

Concept Check

What bound for the tail of a distribution of random variable X uses only expected value of X ?

Markov's inequality.

What condition must be met in order to use Markov's inequality?

The random variable must be non-negative.

True or False: There are distributions for which Chebyshev's inequality is tight (i.e., gives the best possible bound).

True. We gave such a distribution in the lecture.

Lecture 23+24 Summary – Continuous (Two Days)

Two days on continuous probability

Continuous Random Variables

- Idea and Probability Density

- Cumulative Density Function

- Examples

Important Distribution 1: Exponential

Probability Concepts in a Continuous Setting

- Expectation and Variance

- Joint and Marginal Distributions

- Conditional distribution, Independence, and Total Probability

Important Distribution 2: Normal (Gaussian)

- Definition

- Shifting and Scaling

- Standard deviation bands

- Central Limit Theorem

Concept Check

Which continuous distribution is memoryless?

Exponential.

How do you compute the PDF from the CDF?

Take the derivative.

True or False: If X is distributed as a standard normal distribution, then the probability that $-1 \leq X \leq 1$ is approximately 95%.

False. Remember the 68-95-99.7 rule: The probability of being within one standard deviation is $\approx 68\%$. To get 95% the range must span *two* standard deviations of the mean (so $-2 \leq X \leq 2$).

If $X \sim N(45, 10)$ and $\mathbb{P}[X \leq 35] = 0.159$, what is $\mathbb{P}[X \geq 55]$?

Also 0.159 since the distribution is symmetric around $\mu = 45$.

Lecture 25 Summary – Markov Chains I

A Markov Chain is a series of random variables, such that

- We know the distribution of X_0 , denoted as π_0 .
- We know the transition probability matrix P .
- $P(X_{\text{next}} = j | X_{\text{prev}} = i) = P(i, j)$
- The distribution of k th random variable in the chain $\pi_k = \pi_0 P^k$

“First step analysis” problems

- Computations are often easier than other “lower level” techniques.
- Examples:
 - Finding the expected time to reach a given state (getting two tails in a row).
 - Finding the probability of reaching one state before another (making \$100 before running out of money).

Concept Check

Row or Column: The sum of the values in a _____ of the transition matrix must be 1.

Row. Conditional probabilities $P(X_{\text{next}} = j | X_{\text{prev}} = i) = P(i, j)$ must add up to 1 for a fixed condition, and “a fixed condition” here means for a specific value of i (i.e., across a row).

True or False: To determine the probability distribution of state X_{t+1} you must know the history of all previous states $(X_1, \dots, X_t)$.

False. You only need to know the one previous state, X_t .

If “next step equations” look like $f_1(x) = 1 + c_1 f_1(x) + c_2 f_2(x) + \dots$, are the equations for “hitting time” or the “ A before B ” problem?

Equations for the hitting time have the $1 +$ part, since you’re counting steps from the next state (and have one more step than from the following state). The equations for A before B give probabilities, so “ $1 +$ ” doesn’t make sense.

Lecture 26 Summary – Markov Chains II

The Stationary Distribution – two central questions:

- 1 Is there a unique stationary distribution?

Yes if and only if the Markov Chain is irreducible

- 2 Do state probabilities converge to the stationary from any initial distribution?

*If and only if the Markov Chain is irreducible and aperiodic (period 1)
Known as the “Fundamental Theorem of Markov Chains”*

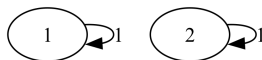
Application: Predicting Text

Idea: Use Markov Chain to predict (or generate!) text

Context is state for Markov chain, and transitions are for next symbol/token

Concept Check

Does the following Markov chain have a stationary distribution?

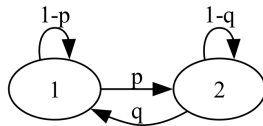


Yes.

Does it have a unique stationary distribution?

No. In fact, *every* distribution is stationary.

Does the following Markov chain converge to a unique stationary distribution for every initial distribution?



Yes, as long as p and q are strictly between 0 and 1. Then it is aperiodic and irreducible, so has a unique stationary distribution.

Good luck tomorrow!