HKN
CS 70 Final Review

Preetham Gujjula
Andrew Liu
Wesley Hsieh

Topics covered

- Probability
- Random Variables
- Bounds
- Discrete Distributions
- Continuous Distributions
- Inference
- Markov Chains

Modification to Monty Hall

Monty Hall has decided that his game show is too predictable, so he has decided that he will only reveal doors some of the time.

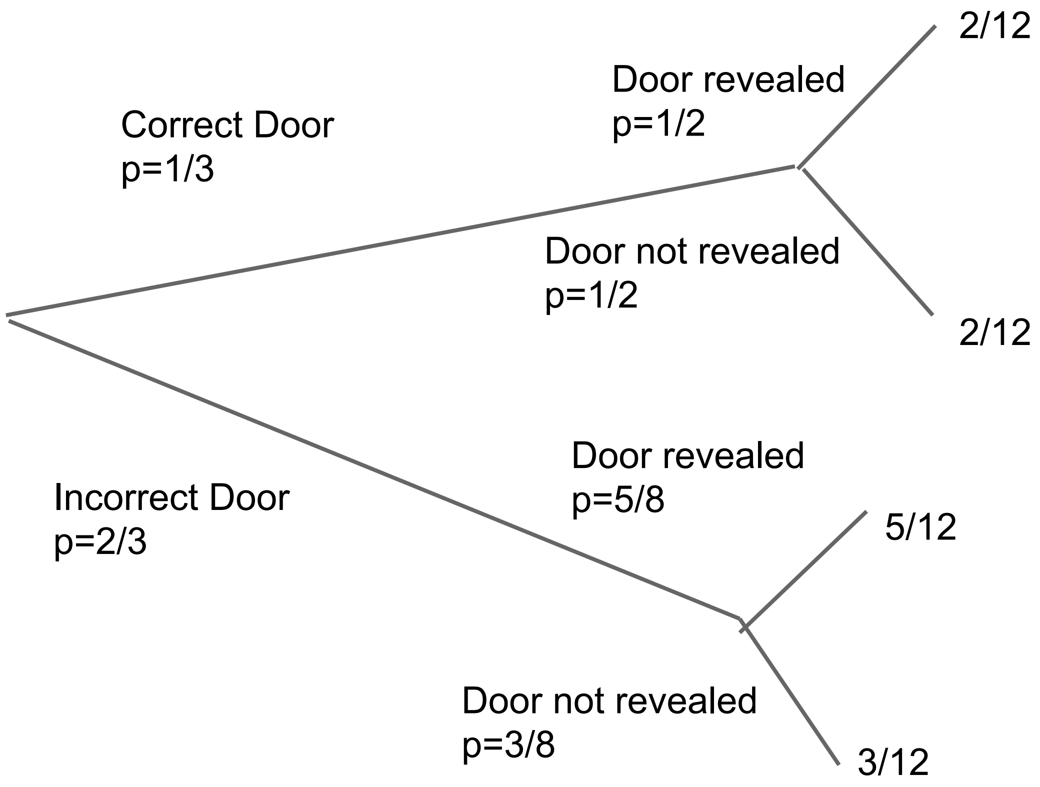
Specifically, if you chose the correct door, he will open a door (revealing a goat) **1/2** of the time. If you chose an incorrect door, he will open a door (revealing a goat) **5/8** of the time.

Should you switch if he opens a door?

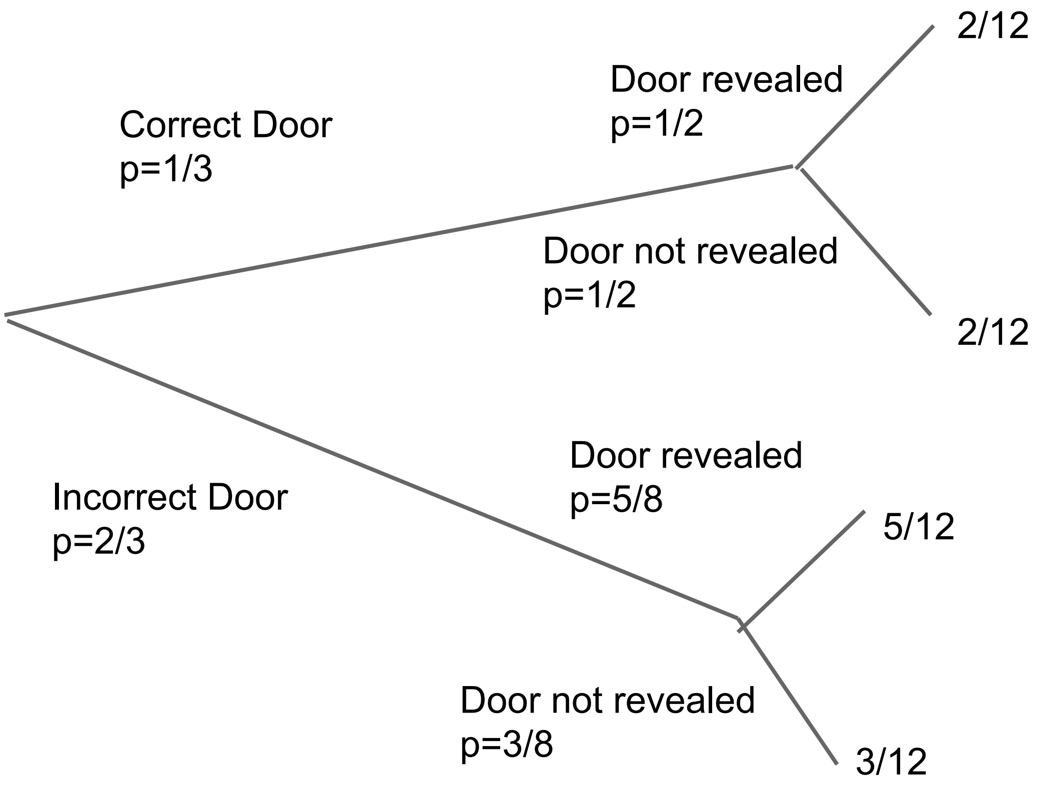
If he doesn't?

Probability Review

Monty Hall Mod Solution



Monty Hall Mod Solution



If a door is opened:

P(correct door) =

$$\frac{2/12}{2/12+5/12} = \frac{2}{7}$$

Switch

If a door isn't opened:

P(correct door) =

$$\frac{2/12}{2/12+3/12} = \frac{2}{5}$$

P(correct door if switch)

(1 - 2/5) / 2 = 3/10

Don't switch

Probabilistic Dating (a)

For this question, there are five men and five women, each with preference lists that have been generated at random.

Richard is trying to figure out Barbara's preferences. He knows she likes him better than she likes Shawn, and that she likes Quincy better than she likes Ulrich. How many different preference lists for Barbara are possible?

Probabilistic Dating (a sol)

Barbara: R > S, Q > U

Preference lists can be grouped into fours, only one of which satisfies the conditions.

5! = 120 total preference lists, so

120/4 = **30 possible preference lists**

Q R S T U
Q S R T U
U R S T Q
U S R T Q

Probabilistic Dating (b)

For this question, there are five men and five women, each with preference lists that have been generated at random.

Shawn, for his part, learns that a consecutive chunk of Candy's preference list is Shawn > Tony > Ulrich. What's the probability he's her first choice?

Probabilistic Dating (b sol)

Candy: Q ? R ? STU

Consider Stu to be one superperson. We're looking for the probability Stu is first in the list. We can see it's ⅓, recalling that because the preferences were random, possible list is equally likely.

Q R S T U
R Q S T U
Q S T U R
R S T U Q
S T U Q R
S T U R Q

Probabilistic Dating (c)

For this question, there are five men and five women, each with preference lists that have been generated at random.

Quincy knows that Diane prefers him to Richard, and that she prefers him to Shawn. What's the probability she likes him best?

Probabilistic Dating (c **sol**)

Diane: Q > R, Q > S

- First, place T and U (5P2 = 20 ways)
- Now, there's only one place for Q.
- Iff neither T nor U was first (4P2 = 12 ways), then Q must be first: 12/20 = **3/5**
- Finally, for each one of these situations, R and S could be in either order. But 24/40 is still 3/5.

Probabilistic Dating (d)

Now, constrain the preference lists as follows:

Male prefs	
q	A B C D E
r	B C D E A
s	C D E A B
t	D E A B C
u	E A B C D

Female prefs		
A	? ? ? ? ?	(A's prefs are random)
B	q r s t u	
C	q r s t u	
D	q r s t u	
E	q r s t u	

If we run P&R algorithm, what's the distribution of Andrea's matchings if men propose? If women propose?

Probabilistic Dating (d sol)

Men-propose

There's only one round, and Andrea is matched with Quincy with 100% probability.

Women-propose

Claim: Andrea can't be stably paired with Richard, Shaun, or Tony.

If Andrea proposes to Ulrich before Quincy, she must end up with Ulrich. Otherwise, she ends up with Quincy (he would never reject her, his first choice). Thus, she is matched with: Quincy with 50% probability, and with Ulrich with 50% probability.

Random Variables, Expectation, Variance

Random Variables

- Random variable X is a representation of an unknown event
- Can take on a variety of values with different probabilities
- Random variables allow us to ask questions about unknown quantities.

X is the number of heads in one coin flip

x	$P(X=x)$
0	0.5
1	0.5

St. Petersburg Lottery

- A casino in St. Petersburg offers the following bet: a fair coin is tossed until a head appears, and the payout, which starts at $1, is doubled for every toss made.
- Payout is 2^t, where t = the total number of tails.
- How much would you pay to enter the bet?
- Calculate the expected value for this bet.

St. Petersburg Lottery

- Calculate the expected value for this bet.

$$
\begin{aligned}
E(B) &= \sum_{a \in \mathcal{A}} a \times Pr[B = a] \\
&= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{8} + 8 \cdot \frac{1}{16} + \cdots \\
&= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots \\
&= \sum_{k=1}^{\infty} \frac{1}{2} = \infty \,.
\end{aligned}
$$

St. Petersburg Lottery

- How can we make sense of this?
- Consider the following modification: the casino no longer has unlimited money. It is backed by the Bill & Melinda Gates Foundation, and only has $\$2^{36}$, or around \$70 billion, in its reserves.
- If the outcome of the bet is higher, the player is given all the money in the reserves, but no more.
- Calculate the new expected value.

St. Petersburg Lottery

$$E(B) = 1 \cdot \frac{1}{2^1} + 2 \cdot \frac{1}{2^2} + 4 \cdot \frac{1}{2^3} + \cdots + 2^{36} \cdot \frac{1}{2^{37}} + 2^{36} \cdot \frac{1}{2^{37}}$$
$$= 38 \cdot \frac{1}{2} = 19$$

Where is this term from?

The infinite expected value from before was due to the infinite extremely high, extremely unlikely terms, most of which we now omit.

Independence? [Fall 2014]

If X and Y are random variables such that E[XY] = E[X]E[Y], are X and Y independent?

Intuition Interlude

- We know the converse is true, but does that tell us anything about this statement?
- Things get a little funky when one of the expectations is 0… Let's try and examine this
- What is the simplest model we can use?

Solution

Solution: The converse is true, but the original statement is not true in general. Consider an example from Note 15, page 10. Let X be a fair coin toss that we consider as taking values $+1$ and -1 equally likely. Suppose Y is an independent fair coin toss that takes values $+1$ and $+2$. The random variables X and Y are independent by construction.

Let's consider a new random variable $Z = XY$. Is Z independent of Y? Obviously not. Z takes on four possible values $-2, -1, +1, +2$ and the magnitude of Z reveals exactly what Y is. We also know that $E[Y] = \frac{1}{2}(1+2) = \frac{3}{2}$, and $E[Z] = \frac{1}{4}(-2-1+1+2) = 0$, so $E[Y]E[Z] = 0$. However,

$$\begin{aligned}
E[YZ] &= \sum_y \sum_z yzPr[Y = y, Z = z] \\
&= \frac{1}{4}(1(-1) + 1(1) + 2(-2) + 2(2)) \\
&= \frac{1}{4}(-1 + 1 - 4 + 4) \\
&= 0,
\end{aligned}$$

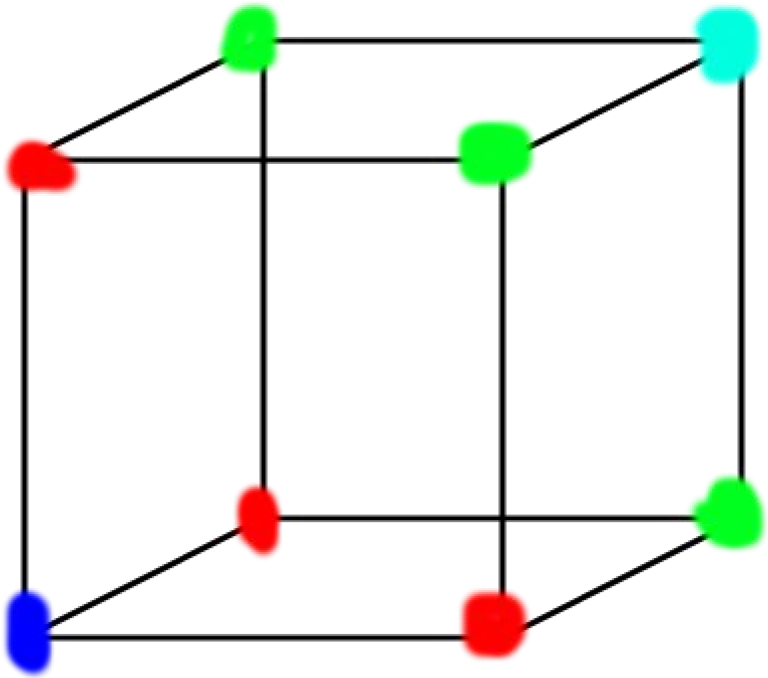
which means $E[YZ] = E[Y]E[Z] = 0$, but Y and Z are not independent. Hence, this is a counterexample to the original claim, and the statement must be false.

Crawling on a Cube

Billy the Bug is on a cube with vertices at $(0, 0, 0), (0, 0, 1), \ldots, (1, 1, 1)$. He starts at the origin and wants to get to $(1, 1, 1)$. However, he has no sense of direction, so at every vertex, he chooses a random edge to travel along (including the one he just came from). What is the expected length of his journey?

Crawling on a Cube

At first glance, it seems like there are 8 (!) state variables to consider. However, note that in the diagram, identically colored vertices all have the same expected number of steps to the finish. Thus, we can consider 4 state variables instead.



Crawling on a Cube

Let a_1, a_2, a_3, and a_4 be the expected number of steps to get to (1, 1, 1) from the blue, red, green, and cyan vertices respectively. Then

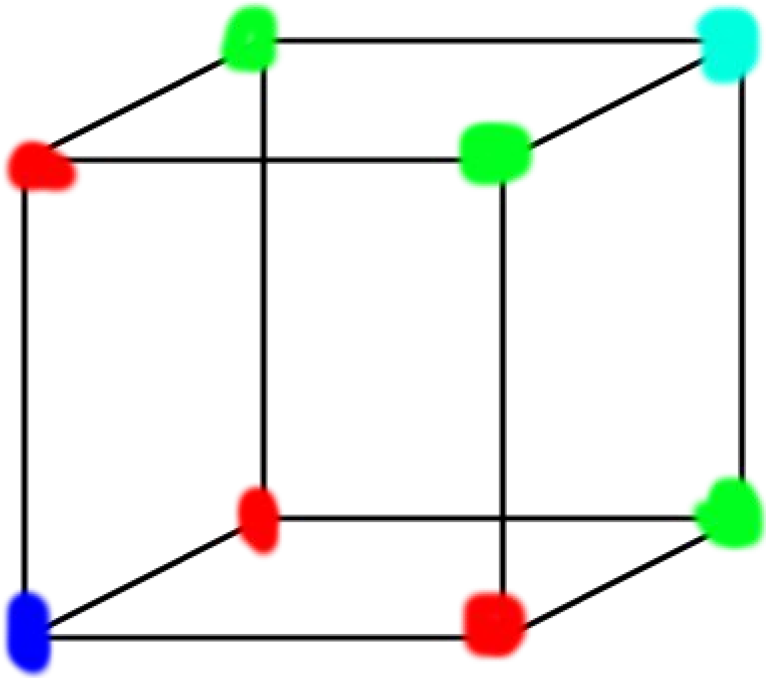
$$a_1 = 1 + a_2$$

$$a_2 = 1 + (1/3)\, a_1 + (2/3)\, a_3$$

$$a_3 = 1 + (1/3)\, a_4 + (2/3)\, a_2$$

$$a_4 = 0$$

Solving these gives $\boldsymbol{a_1 = 10}$.



Pokemon

Saavan is playing Pokemon and wants to “Catch ‘Em All”! To catch Pokemon, he has to go into battle in the wild. Every time he goes into battle, he sees a random Pokemon. If he doesn’t yet own this Pokemon, he catches it and adds it to his collection. If there are n Pokemon, what is the expected number of battles he has to engage in to catch all the Pokemon?

Markov Chains

Markov Review

Measures progression from one state to another state

- Memoryless, your next step is only dependent on your current state
- Given a state, with some probability we move to an adjacent state

Markov Review

Irreducibility

- All states can reach every other state

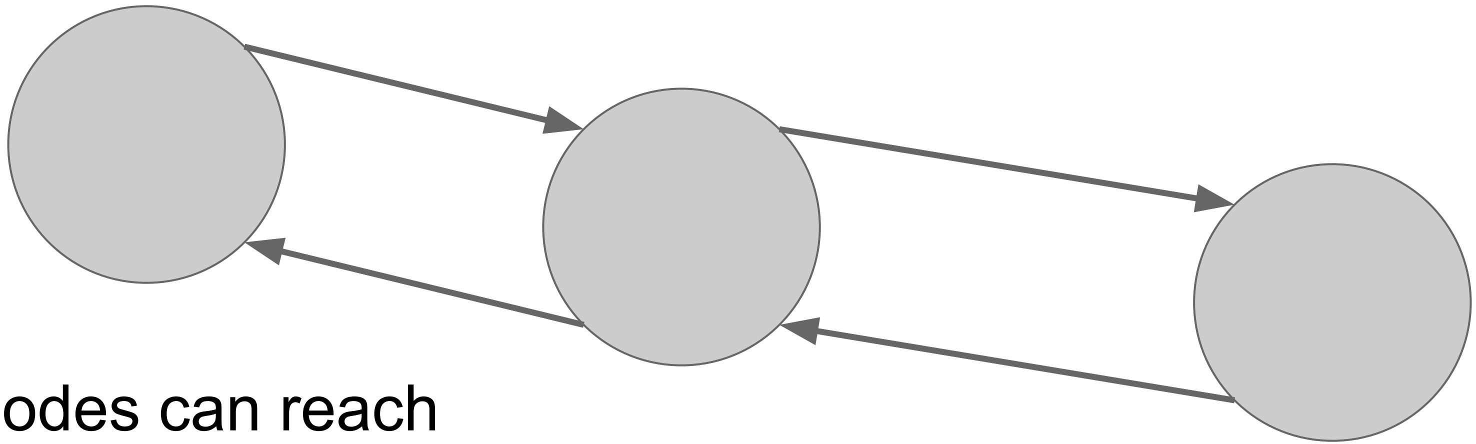
Aperiodic

- Given that you’re at a state, at some point in the future you have non-zero probabilities of being in each other state.
- Look at GCD of self hitting times

Invariant Distribution:

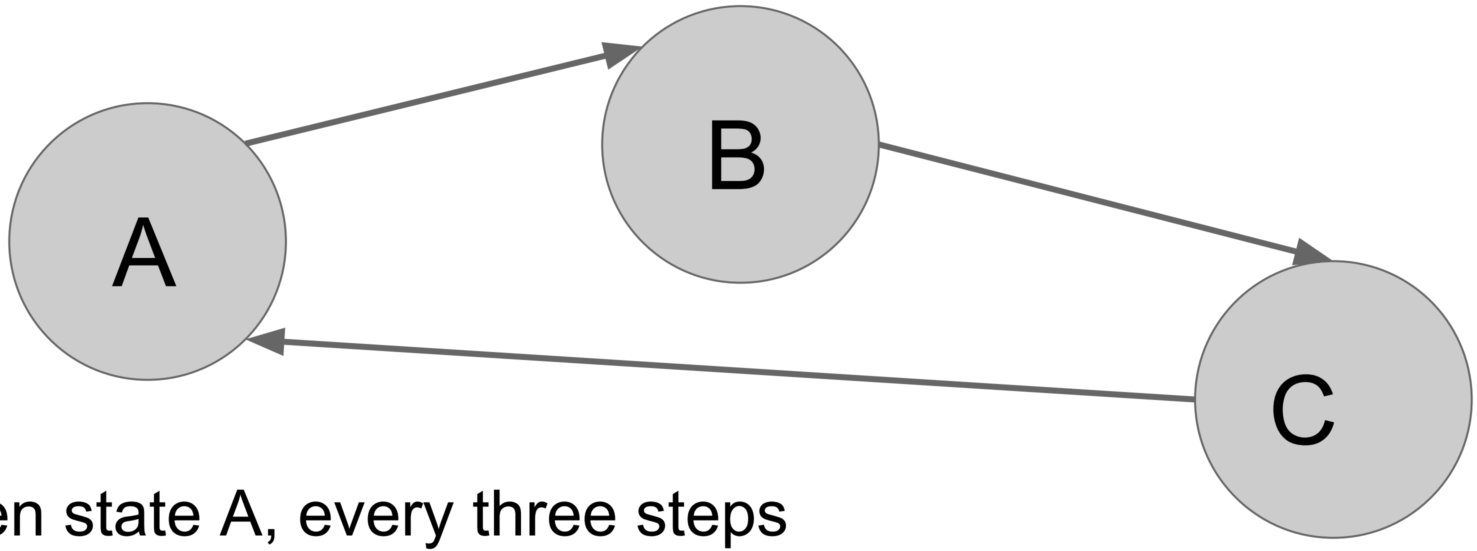
- Steady-state of markov chain
- Not necessarily unique, could depend on your initial distribution

Irreducible Markov



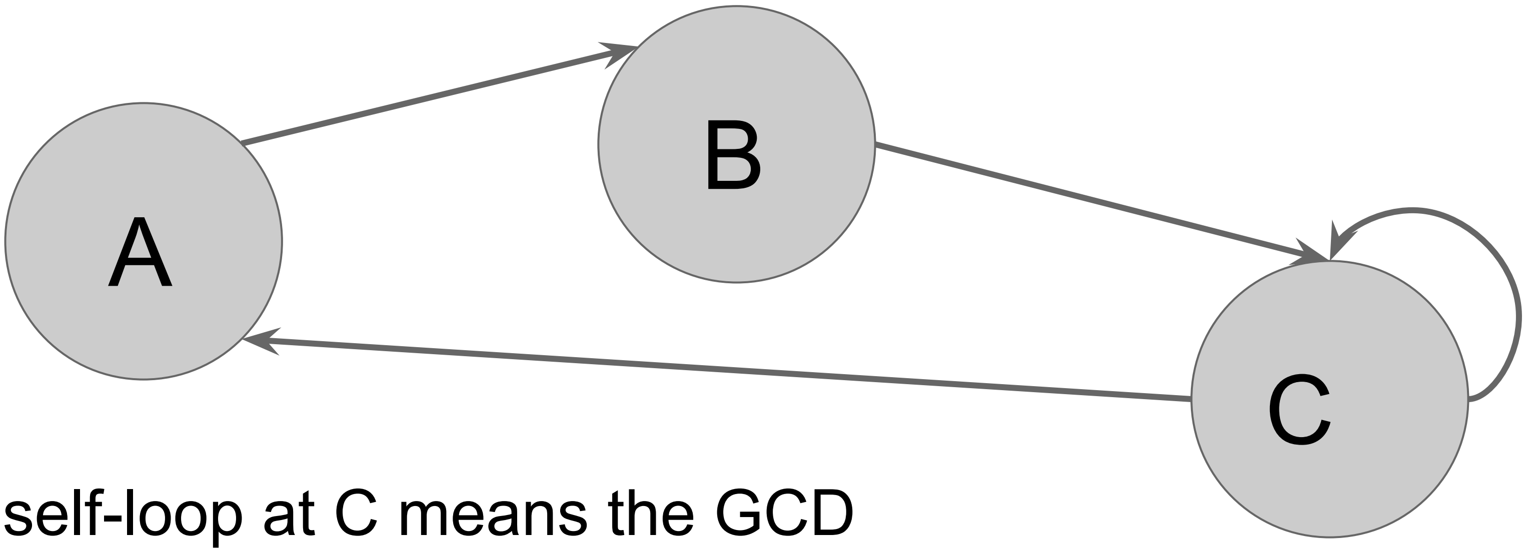
All nodes can reach every other node and vice versa

Periodic Markov



Given state A, every three steps
we end up back at node A.
Therefore the GCD is 3

Aperiodic Markov



The self-loop at C means the GCD is 1

Calculating Hitting Time

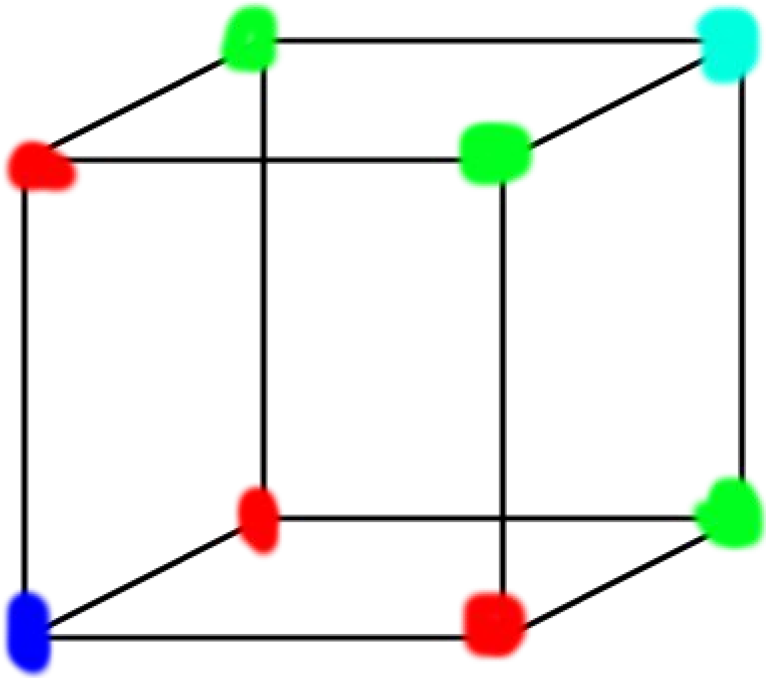
- Crawling on cube was an example of calculating hitting time

Crawling on a Cube

Let a_1, a_2, a_3, and a_4 be the expected number of steps to get to (1, 1, 1) from the blue, red, green, and cyan vertices respectively. Then

$$a_1 = 1 + a_2$$
$$a_2 = 1 + (\tfrac{1}{3})\, a_1 + (\tfrac{2}{3})\, a_3$$
$$a_3 = 1 + (\tfrac{1}{3})\, a_4 + (\tfrac{2}{3})\, a_2$$
$$a_4 = 0$$

Solving these gives $\boldsymbol{a_1 = 10}$.



Calculating Hitting Time

Expected time from a state is (1 + Σ E[S] * Pr(S)) for all adjacent states.

E[Blue] = 1+1*E[Red]

E[Red] = 1+0.33*E[Blue]+0.66*E[Green]

E[Green] = 1+0.66*E[Red]+0.33*E[LB]

E[LB] = 0

Product of Dice

A fair die with labels (1 to 6) is rolled until the product of the last two rolls is 12. What is the expected number of rolls?

Hint: How can we formulate the problem into a Markov Chain?

Product of Dice

We can model this as a markov chain with three states.

1. Start state: We are initially in this state and we reach this state any time we roll a 5 or 1 since those rolls can't make 12. There is a ⅓ probability of staying in this state (by rolling a 5 or 1) and ⅔ probability of reaching the intermediate state (by rolling a 2, 3, 4, or 6)
2. Intermediate state: we are in this state after rolling a 2, 3, 4, or 6. With ⅙ probability we reach the end state (by rolling something that multiplies to 12). With ⅓ probability we reach the start state (by rolling 1 or 5) and with ½ probability we stay in the same state (by rolling a 2, 3, 4, or 6 that doesn't multiply to 12).
3. End state

Compute the hitting time by letting T_i be the hitting time from state i to state 3.

$$T_1 = 1 + \frac{1}{3}T_1 + \frac{2}{3}T_2$$
$$T_2 = 1 + \frac{1}{3}T_1 + \frac{1}{2}T_2,$$

$$T_1 = 10.5 \text{ and } T_2 = 9$$

Bounds

Bound Derivations: Markov

We want to bound $Pr(X \geq \alpha)$ where X is some positive-valued random variable.

Consider the random variable:

$$I = \begin{cases} 1 & \text{if } X \geq \alpha \\ 0 & \text{otherwise} \end{cases}$$

Bound Derivations: Markov

$$I = \begin{cases} 1 & \text{if } X \geq \alpha \\ 0 & \text{otherwise} \end{cases} \longrightarrow I \leq \frac{X}{\alpha}$$

This is easy to see when you enumerate both possibilities: $X \geq \alpha$ and $X < \alpha$

Bound Derivations: Markov

Since expectation is monotonic over positive values, taking the expectation does not change the inequality:

$$I \leq \frac{X}{\alpha} \longrightarrow \mathbb{E}[I] = Pr(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}$$

Bound It *(from Fa14 final)*

A random variable X is always strictly larger than -100. You know $E[X]$ = -60.

Give the best possible upper bound on $P(X \geq -20)$

Bound It solution *(from Fa14 final)*

A random variable X is always strictly larger than -100. You know $E[X] = -60$. Give the best possible upper bound on $P(X \geq -20)$

Since we don't know the variance of X or anything else about its distribution, we'll use a Markov bound.

Since X can be negative, define $Y = X + 100$. Y is positive, and $E[Y] = E[X + 100] = E[X] + 100 = 40$.

$P(X \geq -20) = P(Y \geq 80) \leq E[Y] / 80 = 1/2$

Bound Derivations: Chebyshev

Want to bound $\Pr(|X - \mu| \geq \varepsilon)$

Since squaring is monotonic over positive values this equals $\Pr((X - \mu)^2 \geq \varepsilon^2)$.

Now apply Markov:

$\Pr(|X - \mu| \geq \varepsilon) \leq E[(X - \mu)^2] / \varepsilon^2 = \mathrm{Var}(X) / \varepsilon^2$

All Together Now [Source MIT 6.041]

Let $X_1, \ldots, X_{10}$ be independent random variables, uniformly distributed over the unit interval [0,1].

(a) Estimate $\mathbf{P}(X_1 + \cdots + X_{10} \geq 7)$ using the Markov inequality.

(b) Repeat part (a) using the Chebyshev inequality.

(c) Repeat part (a) using the central limit theorem.

Solution Part a.

To use the Markov inequality, let $X = \sum_{i=1}^{10} X_i$. Then,

$$\mathbf{E}[X] = 10\mathbf{E}[X_i] = 5,$$

and the Markov inequality yields

$$\mathbf{P}(X \geq 7) \leq \frac{5}{7} = 0.7142.$$

Solution Part b.

Using the Chebyshev inequality, we find that

$$\begin{aligned}
2\mathbf{P}(X - 5 \geq 2) &= \mathbf{P}(|X - 5| \geq 2) \\
&\leq \frac{\mathrm{var}(X)}{4} = \frac{10/12}{4} \\
\mathbf{P}(X - 5 \geq 2) &\leq \frac{5}{48} = 0.1042.
\end{aligned}$$

Solution Part c.

Finally, using the Central Limit Theorem, we find that

$$
\begin{aligned}
\mathbf{P}\left(\sum_{i=1}^{10} X_i \geq 7\right) &= 1 - \mathbf{P}\left(\sum_{i=1}^{10} X_i \leq 7\right) \\
&= 1 - \mathbf{P}\left(\frac{\sum_{i=1}^{10} X_i - 5}{\sqrt{10/12}} \leq \frac{7-5}{\sqrt{10/12}}\right) \\
&\approx 1 - \Phi(2.19) \\
&\approx 0.0143.
\end{aligned}
$$

Check

Why does the Central Limit Theorem (CLT) “work better” than Chebyshev?

When does the CLT perform best?

Check

Why does the Central Limit Theorem (CLT) “work better” than Chebyshev?

Its estimate decreases exponentially as the value moves away from the mean.

When does the CLT perform best?

When the deviations from the mean are small (**central** limit theorem)

Discrete Distributions

Binomial Distribution

If X follows *Binom(n,p)*

P(X=x) = (*n* choose *x*)*p*^*x* (1-*p*)^(*n*-*x*)

E[X] = *np* Var[X] = *np*(1-*p*)

Geometric Distribution

If X follows *Geom(p)*

$P(X=x) = (1-p)^x (p)$

$E[X] = 1/p$ $\quad$ $Var[X] = (1-p)/p^2$

Poisson Distribution

If X follows *Poisson*(λ)

P(X=x) = e^(-λ)(λ^x)/(x!)

E[X] = λ Var[X] = λ

Distributions question

Let X,Y~Geom(p)

Compare the likelihood of these events

- P(X=k,Y=k), P(X+Y=2k)
- P(X=k|X+Y=2k), P(X=k,Y=k)
- P(X+Y=2k|X=k), P(X=k)

Distributions answer

- P(X=k,Y=k), P(X+Y=2k)

P(X+Y=2k) = P(X=0,Y=2k) + P(X=1,Y=2k-1) + …

+P(X=k,Y=k) + …

+P(X=2k,Y=0)

So P(X+Y=2k) - P(X=k,Y=k) >= 0

since all terms but P(X=k,Y=k) still exist

and they must be positive

Therefore, P(X=k,Y=k) <= P(X+Y=2k)

Distributions answer

- P(X=k|X+Y=2k), P(X=k,Y=k)

$$P(X = k|X + Y = 2k) = \frac{P(X = k, (X + Y = 2k))}{P(X + Y = 2k)} = \frac{P(X = k, Y = k)}{P(X + Y = 2k)}$$

Since P(X+Y=2k) <= 1,

P(X=k|X+Y=2k) >= P(X=k,Y=k)

Distributions answer

- $P(X+Y=2k|X=k)$, $P(X=k)$

$P(X+Y=2k|X=k) = P(Y=k) = P(X=k)$

Poisson Question

The number of people who go to a market in a day varies according to a poisson process with parameter λ. Each person in the market buys an apple with probability p and a tomato with probability q.

Let the number of apples bought be A, and the number of tomatoes bought be T. Find the following quantities in terms of summations

a. P(A=a)
b. P(A=a|T=t)

Poisson Answer

a. Let N be the number of people. Given N=n, A follows Binom(n, p)

$$P(A=a)=\sum_{n=0}^{\infty}P(N=n)P(A=a|N=n)=\sum_{n=0}^{\infty}\frac{e^{-\lambda}\lambda^n}{n!}\binom{n}{a}p^a(1-p)^{n-a}$$

b.

$$P(A=a|T=t)=\frac{P(A=a,T=t)}{P(T=t)}=\frac{\sum_{n=0}^{\infty}P(N=n)P(A=a,T=t|N=n)}{P(T=t)}=\frac{\sum_{n=0}^{\infty}\frac{e^{-\lambda}\lambda^n}{n!}\binom{n}{a}p^a(1-p)^{n-a}\binom{n}{t}p^t(1-p)^{n-t}}{\sum_{n=0}^{\infty}\frac{e^{-\lambda}\lambda^n}{n!}\binom{n}{t}p^t(1-p)^{n-t}}$$

Continuous Distributions

Taxis and Buses

The taxi and bus stops near Al's home are in the same location. When Al goes there at the same time every morning, a taxi is waiting with probability ⅔. If there is no taxi already there, another taxi will arrive in a time that is uniformly distributed between 0 and 10 and a bus will arrive in 5 minutes. What is Al's expected waiting time?

A Model

We will use the law of total expectation (think the James Bond problem from homework). What events can we condition on? There is an event that there is already a taxi there, call this A; one that the takes takes less than 5 minutes to arrive, call this B; and one that the taxi takes more than 5 minutes to arrive, call this C. Say the random variable representing waiting time is X. Thus we have: $E[X] = P(A)E[X|A] + P(B)E[X|B] + P(C)E[X|C]$

Solution

We note that $P(A) = 2/3$ and $E[X|A] = 0$. In addition, $P(B) = 1/6$ and $E[X|B] = 5/2$. Finally, we have that $P(C) = 1/6$ and $E[X|C] = 5$. Putting this all together gives:

$E[X] = 15/12$.

Exponential

Consider the random variable X that is exponentially distributed with rate 1. Thus: $f_X(x) = e^{-x}$. What is $E[|X - 1|]$?

Calm Before the Storm (Setup)

- How can we make our lives easier when solving this problem?
- Note: exponential random variables are non-negative!
- Can we split up the expectation integral?

Solution (caution: lots of math)

$$
\begin{aligned}
E(|X-1|) &= \int_0^1 (1-x)e^{-x}dx + \int_1^{\infty} (x-1)e^{-x}dx \\
&= \int_0^1 e^{-x}dx - \int_0^1 xe^{-x}dx + \int_1^{\infty} xe^{-x}dx - \int_1^{\infty} e^{-x}dx \\
&= \int_0^1 e^{-x}dx + \int_0^1 xde^{-x} - \int_1^{\infty} xde^{-x} - \int_1^{\infty} e^{-x}dx \\
&= \int_0^1 e^{-x}dx + [xe^{-x}]_0^1 - \int_0^1 e^{-x}dx - [xe^{-x}]_1^{\infty} + \int_1^{\infty} e^{-x}dx - \int_1^{\infty} e^{-x}dx \\
&= [xe^{-x}]_0^1 - [xe^{-x}]_1^{\infty} = 2e^{-1}.
\end{aligned}
$$

Circle Probability

Problem 4. (15 points) A circle of unit radius is thrown on an infinite sheet of graph paper that is grid-ruled with a square grid with squares of unit side. (See Figure 2.) Assume that the center of the circle is uniformly distributed in the square in which it falls. Find the expected number of vertex points of the grid that fall in the circle.

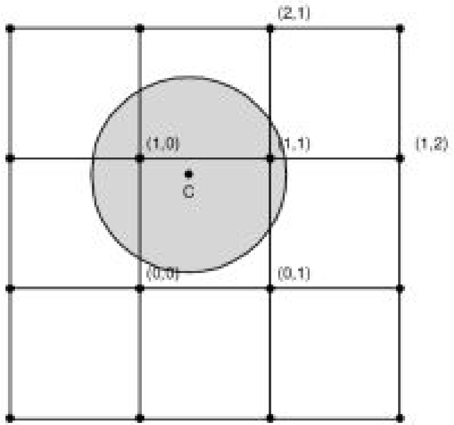


Figure 2: Random circle on a grid-ruled plane.

Problem Breakdown

- This problem seems very hard, but what makes it hard?
- It seems like we need to find probability distributions for a *random* circle to contain 1, 2, 3, or 4 vertices. Wat.
- The problem is asking for an expectation, do you see a way for this to make our lives easier?

Solution Part 1: The First Simplification

- Let's use linearity of expectation!
- Note that the center of the circle falls within some square
- Let X be the number of points within the circle
- Consider indicator random variables X_1, X_2, X_3, X_4 (top left, top right, bottom right, bottom left points respectively) that are 1 if the circle contains the point and 0 otherwise.
- $E[X] = E[X_1+X_2+X_3+X_4] = E[X_1]+E[X_2]+E[X_3]+E[X_4]$

Solution Part 2: The Second Simplification

- Now, what are each of these expectations?
- Draw it! There is a $\frac{\pi}{4}$ area section of the square in which the center can lie to contain a certain vertex (and thus a $\frac{\pi}{4}$ probability of containing said vertex)
- Thus, by linearity, the expected value is π

Sources for Continuous Problems

1. Bertsekas and Tsitsiklis
2. EE126 Fall 2010 Midterm 1
3. EE126 Fall 2014 Midterm 2

Some Final Notes

- A lot of these problems may seem intimidating and foreign, but take them step by step and break them down
- Try to think about the intuition behind a problem before diving head first
- Draw a picture! Often times visualizing what’s going on goes a long way