$P(E_1 \cap E_2)$ 36 3

Quiz 1

2016. 3. 22

1. [2 pts each] Roll two fair dice and let D1 and D2 be the numbers shown on die 1 and 2, respectively. Also define events as follows.

   E1: Event that D1 = 3

   E2: Event that D1 + D2 is even numbers

   E3: D2 >= 4

   a) Compute Pr(E1), Pr(E2) and Pr(E3)

   b) Compute Pr(E1 | E2), Pr(E2 | E1∩E3), and Pr(E1∩E2 | E3)

2. [2 pts] Assume that H1 and H2 are independent events and H2 and H3 are also independent. Then, are H1 and H3 independent also? Prove or give a counter example.

3. [4 pts] Cancer diagnoses problem.

Suspecting that a patient have a cancer, a doctor is performing two examinations. The tests are a) weight loss and b) existence of CEA (carcinoembryonic antigen). Cancer patients suffers from weight loss with probability 0.6 and possess CEA with probability 0.8. The test results of the patient are weight loss and observation of CEA. Assuming that the probability that the patient has cancer before the test, Pr(H), is 0.5. What is the updated probability of the patient having cancer given the test results?

Quiz 2

X^2 $x_i^2 \cdot P(X=x_i)$

$^{k-1} \cdot p$ $1-p$

2016. 3. 31

X^2 p^2

1. (2, 4 pts) Memoryless property of Geometric random variable. Let X be a Geometric random variable with parameter p. Note that we can derive E[X] by conditioning on the result of the first trial.

a) Applying the same approach to derive E[X^2].

b) Let X, and Y be independent Geometric random variables with parameters p and q, respectively. Compute E[max(X, Y)].

2. (2 pts each) Let Xi be a number chosen uniformly at random from [-4, 4].

a) Find Var[Xi]

b) Let X= X1 + X2 +... + Xn. (Xi and Xj are independent) Compute the Markov bound of Pr(X >= n/2).

c) Redo b) for Chebyshev's inequality.

d) Let Y = (Xi)^2.

Write down E[Y] according to the definition of expectation (i.e. $E[Y] = \sum_j y_j \Pr(Y = y_j)$)

And taking this as an example, show that E[Y] can be converted to $E[Y] = \sum_i x_i^2 \Pr(X = x_i)$

<Quiz 2 Solution>

1.

a) (2-p)/(p^2)

E[X^2] = E[X^2 | X1=1]Pr(X1=1) + E[X^2]|X1=0]Pr(X1=0)

= 1*p + E[(X+1)^2](1-p)

= p+(E[X^2]+2E[X]+1)(1-p)

E[X]=1/p

b) 1/p + 1/q - 1/(p+q-pq)

$$E[\max(X,Y)] = \sum_{i\in 0,1, j\in 0,1} E[\max(X,Y)| X_1 = i, Y_1 = j] P(X_1 = i, Y_1 = j)$$

$$= E[\max(X,Y)| X_1 = 0, Y_1 = 0] P(X_1 = 0, Y_1 = 0) + E[\max(X,Y)| X_1 = 1, Y_1 = 0] P(X_1 = 1, Y_1 = 0)$$
$$+ E[\max(X,Y)| X_1 = 0, Y_1 = 1] P(X_1 = 0, Y_1 = 1) + E[\max(X,Y)| X_1 = 1, Y_1 = 1] P(X_1 = 1, Y_1 = 1)$$

$$= E[\max(X,Y)+1](1-p)(1-q) + (E[Y]+1)p(1-q) + (E[X]+1)(1-p)q + pq$$
$$\Leftrightarrow (p+q-pq)E[\max(X,Y)] = 1 + \frac{p}{q} + \frac{q}{p} - q - p$$
$$\Leftrightarrow E[\max(X,Y)] = \frac{1}{p} + \frac{1}{q} - \frac{1}{p+q-pq}$$

2.

a) 20/3

Var[Xi]=(1/9)*(1+4+9+16)*2 = 60/9 = 20/3

E[Xi]=0

c) 40/(3n)

Var[X]=Var[X1+X2+...+Xn] = nVar[Xi]=20n/3

Pr(|X-E[X]|>=n/2) <= Var[X]/(n/2)^2

Pr(|X|>=n/2) <= (20n/3) / (n/2)^2 = 80/(3n)

Pr(X>= n/2)<= 40/(3n)

d)

$$E[Y]$$
$$= 0^2\Pr(Y=0) + 1^2\Pr(Y=1) + 2^2\Pr(Y=4) + 3^2\Pr(Y=9) + 4^2\Pr(Y=16)$$
$$= 0^2\Pr(X=0) + (-1)^2\Pr(X=-1) + 1^2\Pr(X=1) + (-2)^2\Pr(X=-2)$$
$$+ 2^2\Pr(X=2) + (-3)^2\Pr(X=-3) + 3^2\Pr(X=3) + (-4)^2\Pr(X=-4) + 4^2\Pr(X=4)$$

$$= \sum_i x_i^2 \Pr(X = x_i)$$

Quiz 3

2016. 05. 03

1. (5pt. each)

Caution: Detailed analyses of cases and exact counting of number of instances in each case are important. The bounds should be as tight as possible.

Recall the Exercise 6.14.

Given a graph *Gn,p* with p=1/n. X: # triangles in the graph Show that $1/7 \le \Pr(X \ge 1) \le 1/6$ for large n.

Now we modify the problem as follows. Consider a cyclic Hash table with n slots. We put m objects uniformly at random into n slots such that each slot has at most 1 object.

Let X: random variable that counts the number of three consecutive empty slots in the hash table.

Note that if slots 2,3,4,5 are empty while 1 and 6 are not, then there are two instances of three consecutive empty slots (i.e. {2,3,4} and {3,4,5}) and in cyclic Hash table, {6,1,2} is also consecutive slots.

a) When $p=\frac{1}{2}n^{-\frac{1}{3}}$, find the upper bound of $\Pr(X\ge 1)$.

b) When $p=\frac{1}{2}n^{-\frac{1}{3}}$, find the lower bound of $\Pr(X\ge 1)$.

2. Solve Exercise 6.2. (5pt,5pt,10pt.)

A) Prove that for an integer n, there is exists a coloring of the edges of Kn by two colors so that the total # monochromatic K4 is at most *C(n,4) 2^(-5).* B) Give a randomized algorithm for finding a coloring with at most *C(n,4)2^(-5)* monochromatic K4 is that runs in expected time polynomial in n. C) Derandomize the above randomized algorithm using the method of conditional expectation. Your algorithm should be very specific. Just saying that select a color that maximizes (or minimizes) conditional expectation is not enough. (Hint: Order C(n.2) edges. In determining the color of j-th edge, consider the consequences of coloring the edge with one of two colors. Consider all K4 cliques that include the edge. More specifically, if after coloring the edge with one color, some K4 instances becomes (non-)monochromatic, some left with w edges with the same color (that is 6-w edges are not colored yet).

Take Home Quiz (Question 3,4)

Due: Next lecture at 11:00 AM (May 10. Thanks President Park ^^)
You're encouraged to discuss and solve the problems with your colleagues.
But, you should write your own solutions yourself. Please do not share final solutions with your friends.

---

3. We are interested in to prove that p = 1/n is a "threshold" for the property that a Gn,p graph contains a cycle of length 4. (2pt. each)

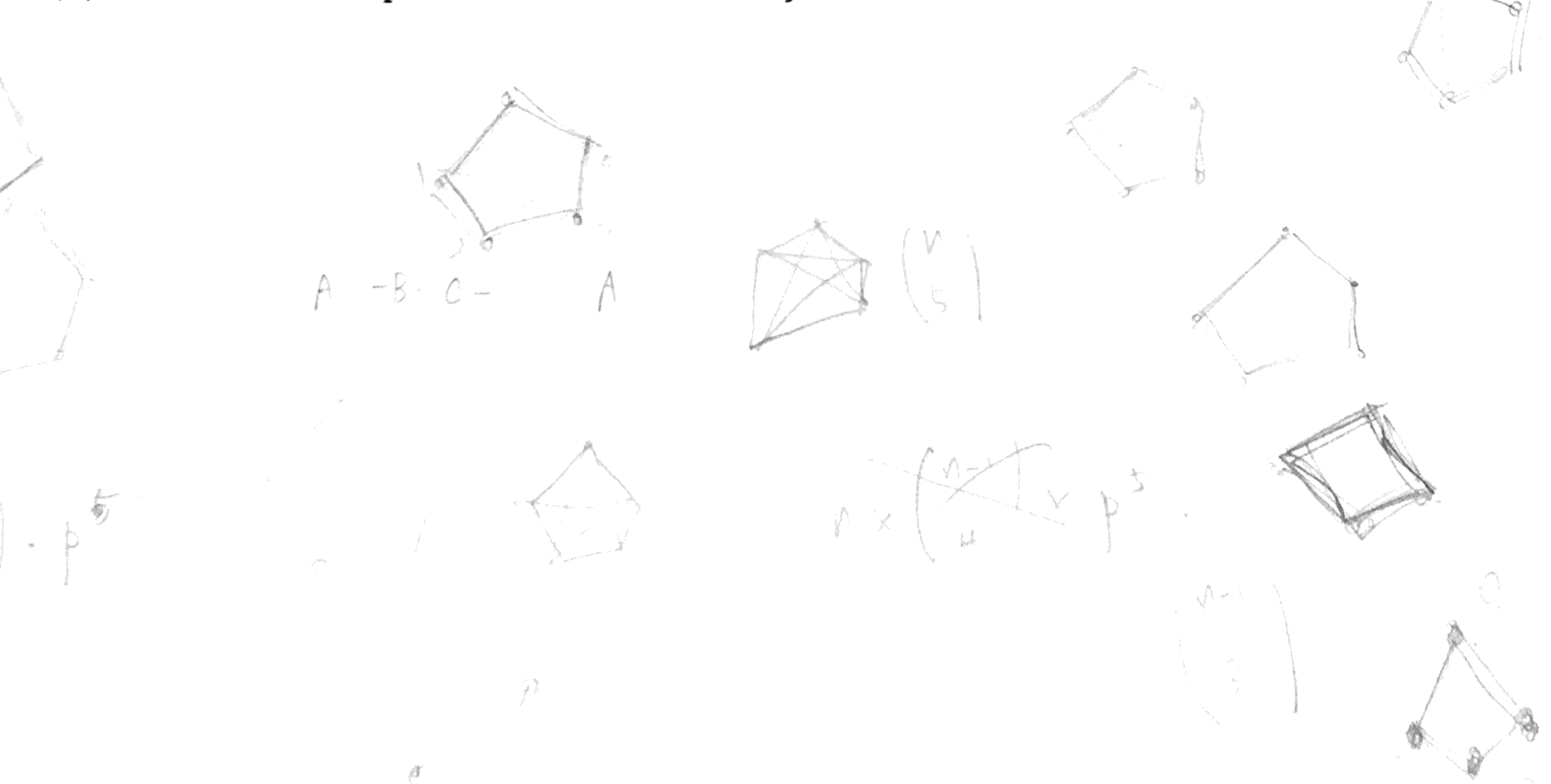
(a) Let the random variable X denote the number of cycles of length 4 in G. Derive E[X] as a function of n and p.
(b) Show that Pr[G contains a cycle of length 4] < ε for p = o(1/n).
(c) Show that Var[X] = $O(n^4p^4 + n^6p^7 + n^5p^6)$.
(d) Show that Pr[No length 4 cycle in G] < ε for p = ω(1/n).

4-1. Prove that, for every integer n, there exists a way to 2-color the edges of K_x so that there is no monochromatic clique of size k when

$$x = n - nCk * 2^{1-kC2}$$

(Hint: Start by 2-coloring the edges of K_n, then fix things up. Sample & Modify is suggested.) (3pt.)

4-2. Let G be a random graph drawn from the $G_{n,1/2}$ model. (2pt. each)
(a) What is the expected number of 5-clique in G?
(b) What is the expected number of 5-cycle in G?



Quiz 4

2016. 5. 17

1. LLL (8pt.)

We learn that the Basic Counting method is used to show a certain properties in <u>edge</u> coloring. Specifically, it was proved the following theorem.

Assume we color edges of K_n graph(A complete graph with n vertices) with one of two colors. **Theorem** If $\binom{n}{k} \cdot 2^{-\binom{k}{2}+1} < 1$, then it is possible to color the edges of K_n such that it has no monochromatic K_k subgraphs

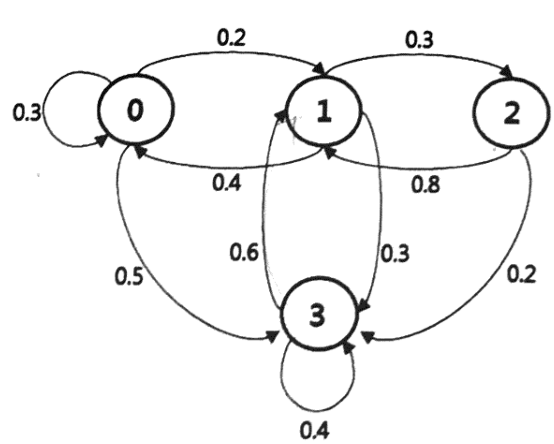
Sketch your strategy to solve the similar edge-coloring property using the LLL. Namely, what are the events (vertices), p? and d? Note that we may change the condition $\binom{n}{k} \cdot 2^{-\binom{k}{2}+1} < 1$

$\binom{n}{k} =$

2. Markov Chain modeling

a) Draw a state-transition diagram of the state transition matrix $P=\begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.9 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}$. (1pt.)

b) What is the state transition matrix corresponding to the state-transition diagram? (1pt.)



c) The original Polya's urn model is that

"There is an urn with m Red and n Blue balls. Randomly select one ball and return the ball with one addition ball with the same color to the urn."

Now we modify the model such that

"There is an urn with m Red and n Blue balls. Randomly select one ball and return the ball with one addition ball with the same color to the urn. After that we delete one ball from the urn such that the total number of balls is m+n"

Let us define the state of the system with the number of Red balls in the urn. Prove that this model is Markov Chain. Draw the state transition diagram of the system. (3pt.)

d) Solve the following problem using Markov Chain.

Hyungho wants to study LLL and Markov Chain, instead of going to see baseball. However, Hyungho cannot cancel his appointment. Therefore, he has to pray for the rainy weather.

Fortunately, He knows the following probabilities.

If the weather has been rainy for the past two days, it will also be rainy tomorrow with probability 0.2.

If the weather is rainy today but was sunny yesterday, it will be rainy tomorrow with probability 0.4.

If it has been rainy for the past two days, it will be rainy tomorrow with probability 0.3.

If it is sunny today but was rainy yesterday, it will be sunny tomorrow with probability 0.5.

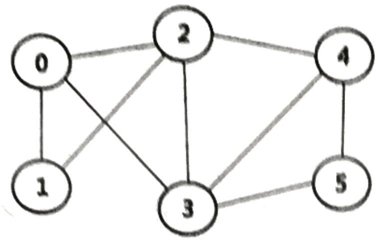
What is the probability of rain in two days later if yesterday and today were sunny? (3pt.)

Quiz 5

2016. 5. 31

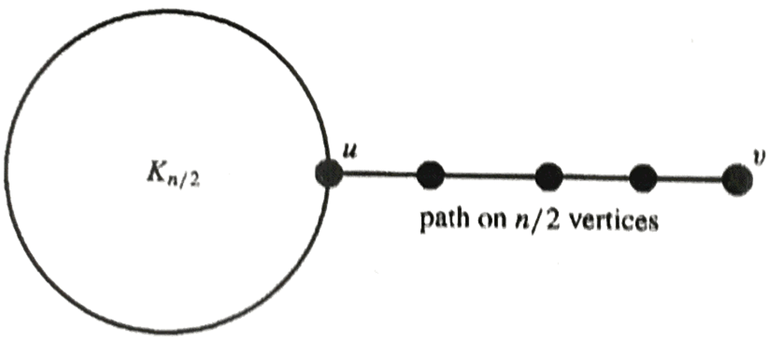
1. 3 pts

Compute the stationary distribution of following random walk graph.



2. 5 pts

The lollipop graph on n vertices is a clique on n/2 vertices connected to a path on n/2 vertices as shown in following graph. The node u is a part of both the clique and the path. Let v denote the other end of the path. The expected covering times of random walks starting at v and u are O(n^2) and O(n^3), respectively. Explain rigorously why the expected covering times are different.



3. 2 pts each

Consider independent random variables $X_1, X_2 \sim \mathrm{Exp}(\theta)$, find the following functions.

a) *Distribution function* $F(X_1, X_2)$

b) *Marginal distribution function* $F_{X_1}(x)$

c) $\Pr(X_1 + X_2 < x | X_1 > y)$

Final Exam

2016. 6. 14.

1. LLL (3 pts each)

a) Dependency Graph

LLL is based on the notion of mutual independency. Prove or give a counter example of the following claim.

Let E1, E2, E3 are pairwise independent events. Then any one event (say E1) is mutually independent of the other two events (Say E2 and E3)?

Definition: Event E1 is mutually independent of events E2, E3,⋯, En if Pr(E1 | E2 ∩E3∩⋯∩En) = Pr(E1)

b) Let $G=(V,E)$ be a Cycle of length $4n$, and let $V=V_1 \cup V_2 \cup V_3 \cup \dots V_n$ be a partition of its $4n$ vertices into n pairwise disjoint subsets, each of cardinality 4. Is it true that there must be an independent set of G containing precisely one vertex from each V_i? (Prove it, or supply a counterexample)

Hint: Let $V_i = \{v_i^{(1)}, v_i^{(2)}, v_i^{(3)}, v_i^{(4)}\}$. Pick a set S by randomly choosing from each V_i one vertex, uniformly and independently.

2. Markov Chain computer systems (3 pts each)

N processes (jobs) are running in a computer system forever. Each of process alternatively uses the CPU and Disk system. Let us say that a process using (or waiting for) CPU/Disk is CPU-hungry and Disk-hungry, respectively. The transition probability from CPU-hungry to Disk-hungry is p and the transition probability from Disk-hungry to CPU-hungry is q. $(0 < p+q \leq 1)$

a) Model the system as a Markov Chain. Define the states and the transition probabilities of the system. Draw the Markov Chain.

b) Compute the stationary distribution. Especially, compute the stationary probability that the CPU is idle (i.e. All processes are Disk-hungry).

c) The performance analysis indicates that, in usual cases, the CPU is the bottleneck. So, we will increase the number of CPU to two. Redraw the Markov Chain with transition probabilities of the improved system.

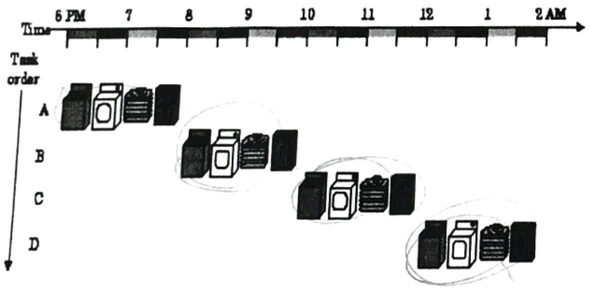
3. **Time reversibility, Burke's Theorem and their application (3 pts each)**

Consider an M/M/1 queueing system where the arrival and departure rates are λ and μ, respectively.

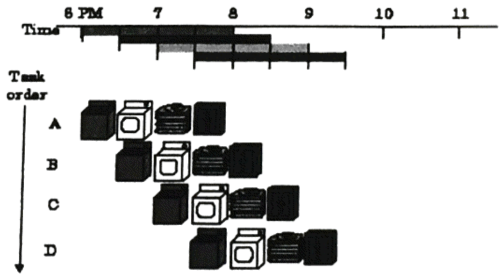
a) Prove that an M/M/1 queueing system is time reversible. (Hint: For the simplicity, consider a DTMC with transition probabilities of λ and μ. Prove that $P(X_m=j \mid X_{m+1}=i, X_{m+2}=i_1, \ldots X_{m+k}=i_k) = \pi_j \ P_{j,i}/\pi_i$

b) According to the Burke's theorem, the departure process of the M/M/1 queue is Poisson with rate λ. Now let us consider a pipelining system consists of two pipe stages. Jobs arrive to the system according to Poisson with rate λ. The service times of the first stage and the second stage have the exponential distribution with means $1/\mu_1$ and $1/\mu_2$, respectively. Compute the number of jobs in the first and the second stages. Also, compute the average time spent in the system (The time from arrival to departure).

c) Now, consider a non-pipelining system. How do you define the service time? (Be as simple as possible. Linearity is good enough.) Compute the average time spent in the system.



Non-pipelining system of 4 stages



Pipelining system of 4 stages

4. **The Monte Carlo Method (2 pts each)**

The problem of counting the number of solutions to a knapsack instance can be defined as follows : Given items with sizes $a_1, a_2, \ldots a_n > 0$ and an integer $b > 0$, find the number of vectors $(x_1, x_2, x_3, \ldots x_n) \in \{0,1\}^n$ such that $\sum_{i=1}^{n} a_i x_i \leq b$. The number b can be thought of as the size of a knapsack, and the x_i denote whether or not each item is put into the knapsack. Counting solutions correspond to counting the number of different sets of items that can be placed in the knapsack without exceeding its capacity.