CS70 – Spring 2024

Lecture 26 – April 23

Today

- Some review problems on Probability & Counting

Next (Last) Lecture

- Applications of Probability in Algorithms

2. Counting [12 points]

Consider the following scenario: There is a test with 30 True/False questions, and each question may either be answered with "T" or "F", or left blank. Solve the following problems. Explanations are **not** required. You should **not** simplify your answers (but your answers should not include summations).

(a) How many possible ways are there to fill out the test? *2pts*

3^{30}

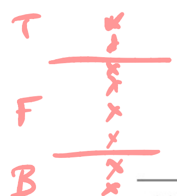
(b) How many ways are there to fill out the test if one answers "T" to exactly 7 questions? *2pts*

$\binom{30}{7} 2^{23}$

(c) How many ways are there to fill out the test if one answers equal numbers of questions with "T", "F" and blank? *2pts*

$$\binom{30}{10}\binom{20}{10} = \frac{30!}{\cancel{20!}\,10!} \times \frac{\cancel{20!}}{10!\,10!} = \frac{30!}{10!\,10!\,10!} = \binom{30}{10\ \ 10\ \ 10}$$

(d) Say one fills out the test so that the first block of questions are all answered with "T", the next block are answered with "F", and the remaining questions are all left blank. How many answer sets are described in this way? *2pts*



$\binom{32}{2}$

(e) How many ways are there to fill out the test leaving exactly 5 questions blank, and answering at least as many questions "T" as "F"? *2pts*

$$\binom{30}{5}\left(\frac{1}{2}2^{25}\right) = \binom{30}{5}2^{24}$$

(f) How many ways are there to fill out the test so that there are exactly 7 "F" answers, no two of which are consecutive? *2pts*

23 non-F answers

$\binom{24}{7} 2^{23}$

Each F answer goes in a slot between the non-F answers $\rightarrow \binom{24}{7}$ ways to position the F answers

Give a combinatorial proof of the following identity:

$$\sum_{i=0}^{n} \binom{n}{i} 2^{n-i} = 3^n, \qquad \text{for all } n \geq 1.$$

RHS = # ways of coloring n items R, B, or G

LHS: first pick i Red items

then color remaining $n-i$ items B or G

5. Quick Bayes.

(5 points) I pick one of two dice with equal probability: one die is a tetrahedron with 1, 2, 3, 4 on the four sides, and one die is a six sided die with 1, 2, 3, 4, 5, 6 on the sides. When rolling a die, all sides are equally likely.

Suppose I roll a 3. What is the probability that I rolled the tetrahedron? (Show work if desired, but clearly indicate final answer.)

T : pick tetrahedron S : six-sided die "3": rolled 3

$$\Pr[3|T] = 1/4 \qquad \Pr[3|S] = 1/6$$

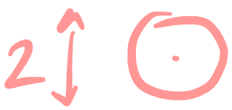
$$\Pr[T] = 1/2 = \Pr[S]$$

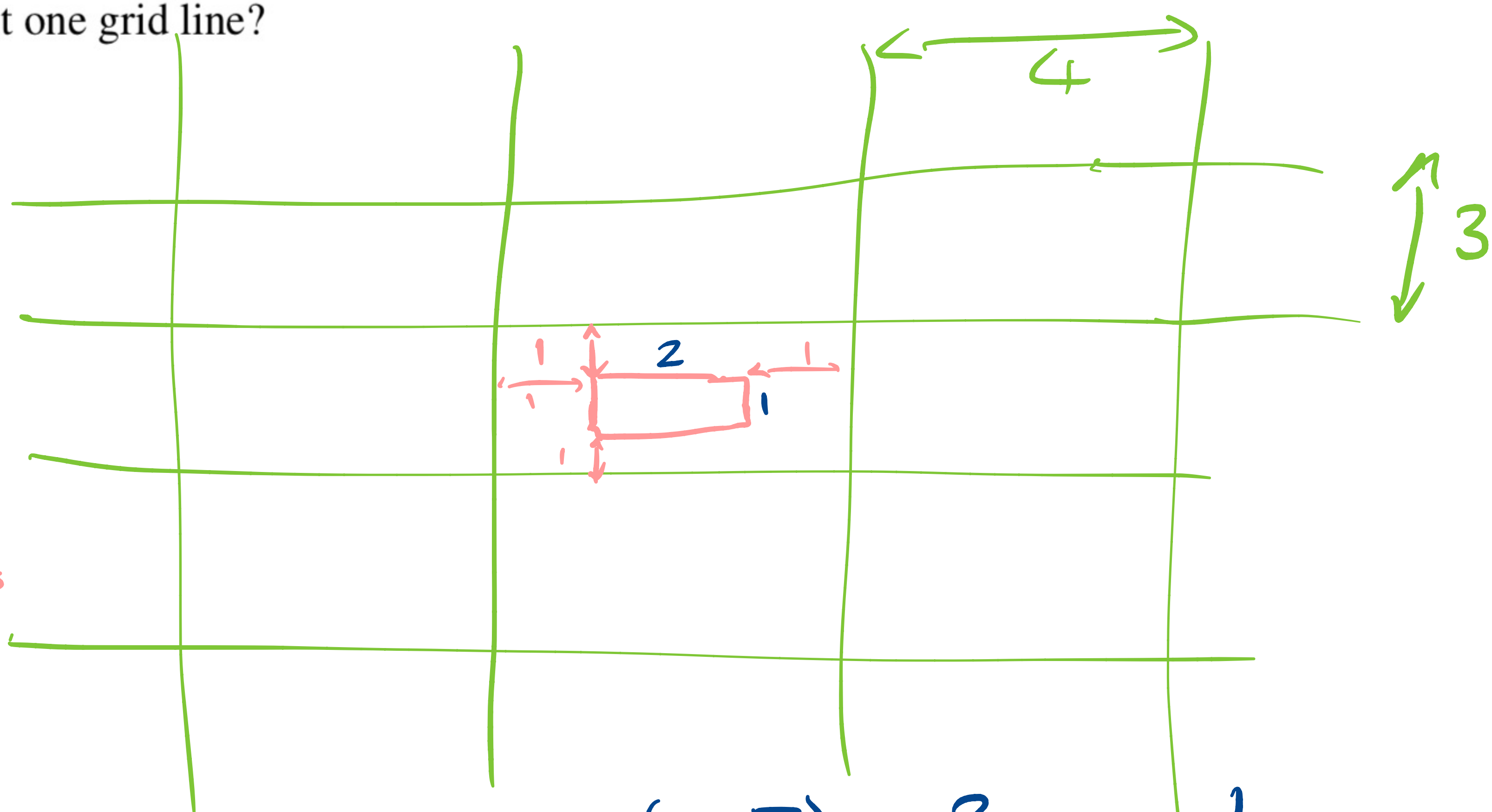
$$\Pr[T|3] = \frac{\Pr[3|T]\Pr[T]}{\Pr[3]} \qquad \text{Bayes Rule}$$

$$= \frac{\Pr[3|T]\Pr[T]}{\Pr[3|T]\Pr[T] + \Pr[3|S]\Pr[S]} \qquad S \equiv \bar{T}$$

$$= \frac{1/4 \times 1/2}{(1/4 \times 1/2) + (1/6 \times 1/2)} = \frac{1/4}{1/4 + 1/6} = \boxed{\frac{3}{5}}$$

Consider randomly dropping a circular coin of radius 1 cm onto a large rectangular grid where horizontal lines are 3 cm apart, while vertical lines are 4 cm apart. What is the probability that the coin intersects at least one grid line?





Don't hit lines
$\Longleftrightarrow$
Center of coin is inside pink rectangle

$$\Pr[\text{don't hit lines}] = \frac{\text{area(pink } \square)}{\text{area(green } \square)} = \frac{2}{12} = \frac{1}{6}$$

$$\Pr[\text{hit a line}] = \boxed{5/6}$$

12. Tails, tails, and tails.

Suppose Shreyas stores some number S initialized to 0. Every time he flips a fair coin, if it lands heads he increments S by 1 and if it lands tails he decrements S by 1. He wishes to calculate $\mathbb{P}[S \geq 20]$ after flipping the coin 100 times.

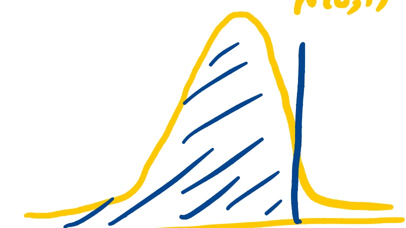
$S = \#H - \#T$ $\quad X = \#H \sim Bin(100, 1/2)$ $\quad S \geq 20$ $\quad \boxed{X \geq 60}$

1. (5 points) Provide an exact answer using a summation. (Hint: $S \geq 20$ is equivalent to saying that Shreyas flips 20 more Heads than Tails.)

$$Pr[S \geq 20] = Pr[X \geq 60] = \sum_{k=60}^{100} \binom{100}{k} \cancel{p^k(1-p)^{n-k}} \left(\tfrac{1}{2}\right)^n$$

Standard Normal

$N(0,1)$



$$\Phi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$$

2. (5 points) Provide an upper bound using Markov's inequality.

$$Pr[X \geq 60] \leq \frac{E[X]}{60} = \frac{50}{60} = \boxed{5/6}$$

$$Var(X) = n \cdot p \cdot (1-p) = 100 \times \tfrac{1}{2} \times \tfrac{1}{2} = 25$$

3. (5 points) Provide an upper bound using Chebyshev's inequality.

$$Pr[X \geq 60] = Pr[X - \underbrace{E[X]}_{=50} \geq 10]$$

$$\leq Pr[|X - E[X]| \geq 10] \leq \frac{Var(X)}{(10)^2} = \frac{25}{100} = \boxed{\frac{1}{4}}$$

4. (5 points) Provide an upper bound using the Central Limit Theorem. (Leave your answer in terms of Φ, where Φ is the standard Normal CDF, if you deem necessary.)

$$Z = \frac{X - E[X]}{\sigma(X)} = \frac{X - 50}{5} \sim N(0,1)$$

$$Pr[X \geq 60] = Pr\left[Z \geq \frac{60-50}{5}\right] = Pr[Z \geq 2] = \boxed{1 - \Phi(2)} \approx 0.02$$

You are given an urn containing N balls, r of which are red and $N-r$ blue. Suppose you pick k balls at random from the urn *with* replacement, and let R be the number of red balls in your sample. Define the r.v. $P = \frac{R}{k}$.

(a) What are $\mathbb{E}[P]$ and $\mathrm{Var}(P)$?

$P = \frac{1}{k} \times \mathrm{Bin}(k, \frac{r}{N})$

$E[P] = \frac{1}{k} \times \frac{kr}{N} = \boxed{\frac{r}{N}}$

$\mathrm{Var}(P) = \frac{1}{k^2} \times k \times \frac{r}{N} \times (1-\frac{r}{N}) = \boxed{\frac{1}{k} \frac{r}{N} \frac{N-r}{N}}$

(b) Now suppose instead that you sample k balls *without* replacement (where $k \leq N$), and define the r.v.'s R and P as above. What are now $\mathbb{E}[P]$ and $\mathrm{Var}(P)$?

$P = \frac{1}{k}(y_1 + \cdots + y_k)$ $\quad y_i = \begin{cases} 1 & \text{if ith ball is Red} \\ 0 & \text{o.w.} \end{cases}$

NOTE: $E[y_i] = E[y_1] = \frac{r}{N}$ and $E[y_i y_j] = \Pr[y_i = y_j = 1] = \Pr[y_1 = y_2 = 1] = \frac{r}{N} \times \frac{r-1}{N-1}$

(by symmetry)

$E[P] = \frac{1}{k} \sum E[y_i] = \frac{r}{N}$.

$\mathrm{Var}(P) = \frac{1}{k^2} E[(\sum y_i)^2] - E[P]^2$.

$E[(\sum y_i)^2] = \sum_i E[y_i^2] + \sum_{i \neq j} E[y_i y_j] = \frac{kr}{N} + k(k-1)\frac{r}{N} \cdot \frac{r-1}{N-1} \Rightarrow \mathrm{Var}(P) = \frac{1}{k^2}\left\{\frac{kr}{N} + \frac{k(k-1) r}{N} \cdot \frac{r-1}{N-1} - \left(\frac{kr}{N}\right)^2\right\}$

$= \frac{r}{kN}\left\{1 + \frac{(k-1)(r-1)}{N-1} - \frac{kr}{N}\right\}$

$= \boxed{\frac{1}{k} \cdot \frac{r}{N} \cdot \frac{N-r}{N} \cdot \frac{N-k}{N-1}}$

(c) Finally, suppose you are using Chebyshev's inequality to determine the sample size k required to compute an estimate of the proportion $\frac{r}{N}$ of red balls in the urn within specified accuracy and confidence. How does the sample size change when you switch from sampling with replacement to sampling without replacement?

Recall that sample size $\propto \mathrm{Var}(P)$ (for fixed accuracy/confidence and fixed $E[P]$)

Difference in Variances between parts (a) & (b) is a factor $\frac{N-k}{N-1} \leq 1$. So sample size <u>decreases</u> by factor $\frac{N-k}{N-1}$ when sampling w/o replacement.

(a) What distribution would you use to model the following random variables? Write the parameters where possible.

(i) The number of soldiers killed due to horse kicks in the Prussian cavalry in a month, given that the average number of such deaths *per year* is 9 and the total size of the cavalry is large.

Poisson $(3/4)$

(ii) The time between two successive soldiers' deaths in the cavalry due to horse kicks.

Exp $(3/4)$

(iii) The number of black horses in a regiment of 100, given that 20% of the horses in the entire cavalry are black.

Bin $(100, 1/5)$

(iv) The length of a horse's tail measured in centimetres.

$N(??, ??)$

(v) The number of times a new soldier rides his horse before he falls off, given that on each ride he falls off with probability 0.1.

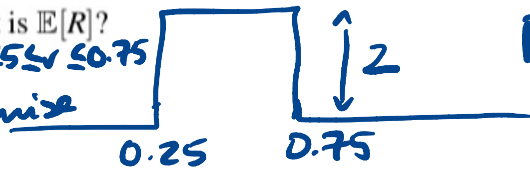
Geom $(1/10)$

3. Uniform parameter for a geometric distribution.

Suppose Jonathan picks a real number R uniformly at random from the range $[0.25, 0.75]$. Then, he takes a coin that yields heads with probability R and tails otherwise and flips it until it yields heads. Let J denote the number of flips (including the last flip that yields heads).

1. What is $\mathbb{E}[R]$?

$$f(r) = \begin{cases} 2 & 0.25 \le r \le 0.75 \\ 0 & \text{otherwise} \end{cases}$$



pdf of R

$E[R] = 0.5$ by symmetry OR:

$$E[R] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_{0.25}^{0.75} 2x\,dx = x^2 \Big|_{0.25}^{0.75} = \boxed{1/2}.$$

2. (5 points) Compute $\mathbb{E}[J]$. Show your work and clearly indicate your final answer.

$$E[J] = \int_{-\infty}^{\infty} \underbrace{E[J \mid R=r]}_{1/r} f(r)\,dr$$

$$= \int_{0.25}^{0.75} \frac{1}{r} \cdot 2 \cdot dr = 2\ln r \Big|_{0.25}^{0.75}$$

$$= \boxed{2\ln 3}$$

3. Compute $\mathbb{P}[J > 70 \mid R = r]$.

$$\Pr[J > 70 \mid R = r] = \Pr[\mathrm{Geom}(r) > 70] = \boxed{(1-r)^{70}}$$

4. (5 points) Compute $\mathbb{P}[J > 70]$. Show your work and clearly indicate your final answer. (The answer is a bit messy.)

Law of Total Prob:

$$\Pr[J > 70] = \int_{-\infty}^{\infty} \Pr[J > 70 \mid R = r] f(r)\,dr = \int_{0.25}^{0.75} (1-r)^{70} \cdot 2\,dr$$

$$= -\frac{2}{71}(1-r)^{71} \Big|_{0.25}^{0.75} = \boxed{\frac{2}{71}\left((0.75)^{71} - (0.25)^{71}\right)}$$

14. How many? And for how long?

(6 points) You purchase a box of light bulbs where the number of light bulbs is a Poisson random variable, $N \sim \text{Poisson}(\mu)$.

Let the random variable X_i denote the lifetime of the i^{th} bulb. The lifetimes of the bulbs are independent.

Moreover, each X_i obeys the exponential distribution

$$f_{X_i}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0, \end{cases}$$

where $\lambda > 0$.

Determine a reasonably-simple expression for $\mathbb{E}[T]$, the expected total life that you can get from the light bulbs in your box. Your expression must be in closed form, and involve nothing other than a subset of the parameters λ and μ.

$$T = X_1 + X_2 + \cdots + X_N$$

$N \sim \text{Poisson}(\mu)$ $\quad E[N] = \mu$

$X_i \sim \text{Exp}(\lambda)$ $\quad E[X_i] = 1/\lambda$

$$E[T/N=k] = E[X_1 + \cdots + X_k] = k/\lambda$$

$$E[T] = \sum_{k=0}^{\infty} E[T/N=k]\Pr[N=k] = \sum_{k=0}^{\infty} \frac{k}{\lambda} \cdot \Pr[N=k] \quad = e^{-\lambda}\frac{\lambda^k}{k!}$$

"Total Expectation"

$$= \frac{1}{\lambda} \sum_{k=0}^{\infty} k \Pr[N=k]$$

$$= \frac{1}{\lambda} E[N] = \boxed{\mu/\lambda}$$

Suppose $X \sim \text{Normal}(1, 2)$ and $Y \sim \text{Normal}(2, 1)$ are independent random variables. What is the distribution of $2X - Y + 1$? State its name and specify its parameter(s).

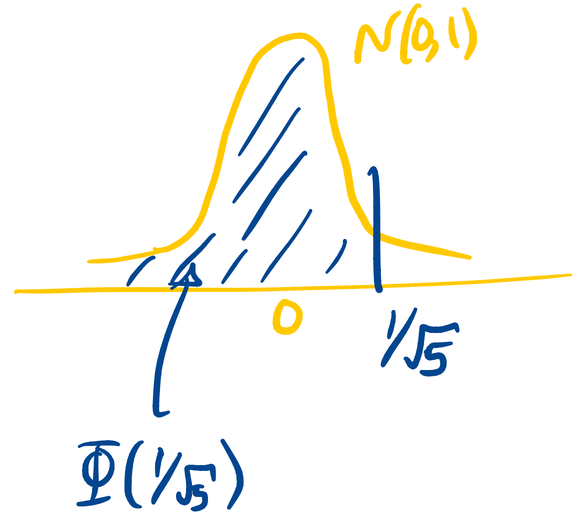
$$N(1, 9)$$

$$E[2X-Y+1] = 2E[X] - E[Y] + 1 = 2 - 2 + 1 = 1$$
$$Var(2X-Y+1) = 4Var(X) + Var(Y) = (4 \times 2) + 1 = 9$$

Suppose A and B are independent $\text{Normal}(1, 1)$ random variables. Find $\mathbb{P}[2A + B \geq 4]$ in terms of the cumulative distribution function (c.d.f.) Φ of the standard normal distribution.

$$Z = 2A + B \sim N(3, 5) \implies \frac{Z-3}{\sqrt{5}} \sim N(0,1)$$

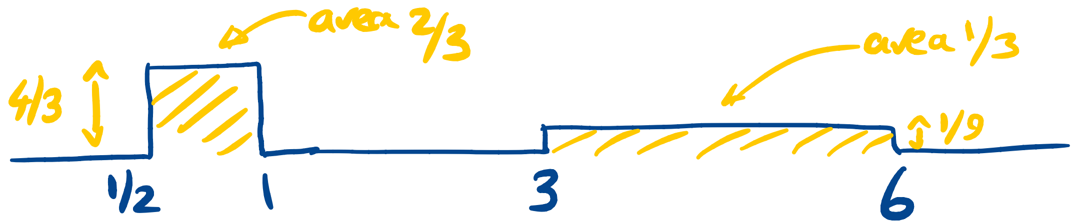
$$Pr[Z \geq 4] = Pr\left[\frac{Z-3}{\sqrt{5}} \geq \frac{1}{\sqrt{5}}\right] = 1 - \Phi\left(\frac{1}{\sqrt{5}}\right)$$



8. Love Thy Neighbor [12 points]

Every morning, my neighbor's dog barks either for a short time or for a long time. The probability he will bark for a short time is 2/3 and the probability he will bark for a long time is 1/3. Given that he barks for a short time, the actual length of his bark (in minutes) is distributed uniformly in the range $[0.5, 1.0]$; the length of a long bark is distributed uniformly in the range $[3.0, 6.0]$. Answer the following questions.

(a) Let the random variable T be the length of the dog's bark (in minutes) on any given morning. Sketch the probability density function of T. [NOTE: Be sure to label all relevant heights etc. on your graph.]



(b) What is $\Pr[T \geq 5]$?

$$\Pr[T \geq 5] = \int_5^\infty f(x)\,dx = \int_5^6 \frac{1}{9}\,dx = \boxed{1/9}$$

(c) What is the expectation $E(T)$? Show your working.

$$E[T] = E[T \mid \text{short}]\Pr[\text{short}] + E[T \mid \text{long}]\Pr[\text{long}]$$
$$= \frac{3}{4} \times \frac{2}{3} + \frac{9}{2} \times \frac{1}{3} = \boxed{2}$$

(Can also do by integration!)

(d) My neighbor receives a Fedex delivery each morning with probability 1/10. Moreover, on days when he receives a delivery, the probability that the dog barks for a long time increases to 4/5 (and the probability of a short bark decreases to 1/5). Now suppose that, on a given morning, I hear the dog barking for a long time. What is my best guess for the probability that my neighbor has received a delivery that day? Show your working.

L = long bark
F = Fedex delivery

Know: $\Pr[L \mid F] = 4/5 \quad \Pr[L \mid \bar{F}] = 1/3 \quad \Pr[F] = 1/10$

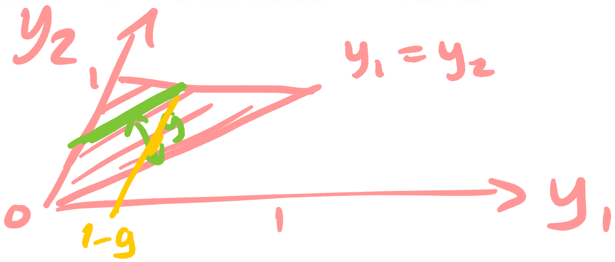
Want: $$\Pr[F \mid L] = \frac{\Pr[L \mid F]\Pr[F]}{\Pr[L \mid F]\Pr[F] + \Pr[L \mid \bar{F}]\Pr[\bar{F}]} = \frac{4/5 \times 1/10}{(4/5 \times 1/10) + (1/3 \times 9/10)}$$

Bayes

$$= \frac{4/5}{4/5 + 3} = \boxed{\frac{4}{19}}$$

Let X_1, X_2 be independent, continuous Uniform$[0,1]$ random variables, and let $Y_1 = \min\{X_1, X_2\}$, $Y_2 = \max\{X_1, X_2\}$.

(a) Find the joint pdf $f(y_1, y_2)$ of Y_1, Y_2.



$$f(y_1, y_2) = \begin{cases} 2 & \text{on } \triangle \\ 0 & \text{otherwise} \end{cases}$$

$\boxed{\begin{array}{l} f \text{ is non-zero only on the } \triangle \\ f \text{ is uniform on the } \triangle \end{array}}$

(b) Let $G = Y_2 - Y_1$ denote the "gap" between Y_1, Y_2. Find the pdf of G.

$$f_G(g) = \int_0^1 f(y_1, y_1+g)\,dy_1 = \int_0^{1-g} 2\,dy_1$$

$$= \boxed{2(1-g) \text{ for } 0 \le g \le 1}$$

$$\Pr[G \in (g, g+\delta g)]$$

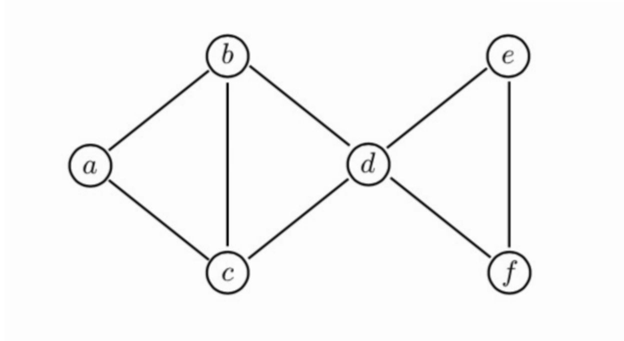
$$= \sum_{y_1} \Pr[Y_1 \in (y_1, y_1+\delta y_1), Y_2 \in (y_1+g, y_1+g+\delta g)]\,\delta y_1 \delta g$$

(c) Compute $\Pr[G \ge \frac{1}{2}]$.

$$\Pr[G \geq 1/2] = \int_{1/2}^1 f_G(g)\,dg = \int_{1/2}^1 2(1-g)\,dg$$

$$= -(1-g)^2 \Big|_{1/2}^1 = \boxed{\frac{1}{4}}$$

Consider random walk on the following directed graph.



(a) If you observe the walk after many steps, starting from vertex a, what is the probability of finding it at vertex d?

Random walk is irreducible (graph is connected) & aperiodic (graph has triangle so is not bipartite).

Hence r.w. converges to unique invariant dist. $\pi(v) = \dfrac{\deg(v)}{2|E|}$ starting from any vertex

And $\pi(d) = \deg(d)/16 = 4/16 = \boxed{1/4}$

(b) What is the expected number of steps until the walk reaches vertex d starting from vertex a?

Let $\beta(i) = E[\#\text{ steps to reach } d \text{ from } i]$. $\beta(d) = 0$. want $\beta(a)$. Note: $\beta(b) = \beta(c)$. Also $\beta(e), \beta(f)$ are irrelevant because we reach d before e, f

$\beta(a) = 1 + \frac{1}{2}(\beta(b) + \beta(c)) = 1 + \beta(b) \Rightarrow \beta(a) = 1 + \beta(b)$

$\beta(b) = 1 + \frac{1}{3}(\beta(a) + \beta(c) + \beta(d)) = 1 + \frac{1}{3}\beta(a) + \frac{1}{3}\beta(b) \Rightarrow 2\beta(b) = 3 + \beta(a)$

Solve green equns: $\beta(a) = 2\beta(b) - 3 = 2(\beta(a) - 1) - 3 = 2\beta(a) - 5$

$\Rightarrow \boxed{\beta(a) = 5}$

(c) Compute the probability that the walk, starting at vertex d, hits vertex a before vertex e.

Let $\alpha(i) = \Pr[\text{hit } a \text{ before } e, \text{ starting at } i]$.

$\alpha(a) = 1$; $\alpha(e) = 0$; $\alpha(b) = \alpha(c)$ (symmetry) Notation: replace $\alpha(i)$ by i !

$b = \frac{1}{3}(a + c + d) = \frac{1}{3} + \frac{1}{3}b + \frac{1}{3}d \Rightarrow 2b = 1 + d$

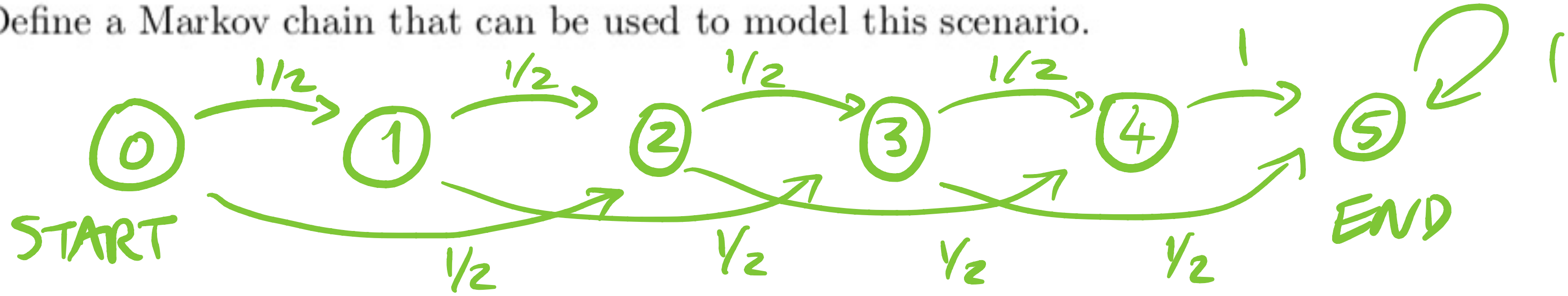
$d = \frac{1}{4}(b + c + e + f) = \frac{1}{2}b + \frac{1}{4}f \Rightarrow 4d = 2b + f$

$f = \frac{1}{2}(e + d) = \frac{1}{2}d \Rightarrow 2f = d$

Solve: $\boxed{\alpha(d) = 2/5}$

Sophie wants to cross a pond using a sequence of four stepping stones, labeled 1, 2, 3, 4. She starts on one edge of the pond (labeled 0) and needs to end up on the other edge (labeled 5). Each five seconds, with probability $\frac{1}{2}$ she proceeds to the next stone, and with probability $\frac{1}{2}$ she jumps two stones ahead (except that from stone 4 she always proceeds to the final edge of the pond).

(a) Define a Markov chain that can be used to model this scenario.



(b) Compute the expected time (in seconds) for Sophie to cross the pond.

$\beta(i) = E[\text{\# steps to reach 5 starting from } i]$

$\beta(5) = 0 \qquad \beta(4) = 1 + \beta(5) = 1 \qquad \underline{\text{Want}}: \beta(0)$

$\beta(3) = 1 + \frac{1}{2}\beta(4) + \frac{1}{2}\beta(5) = 1 + \frac{1}{2} + 0 = 3/2$

$\beta(2) = 1 + \frac{1}{2}\beta(3) + \frac{1}{2}\beta(4) = 1 + \frac{1}{2}\left(\frac{3}{2} + 1\right) = 9/4$

$\beta(1) = 1 + \frac{1}{2}\beta(2) + \frac{1}{2}\beta(3) = 1 + \frac{1}{2}\left(\frac{9}{4} + \frac{3}{2}\right) = 23/8$

$\beta(0) = 1 + \frac{1}{2}\beta(1) + \frac{1}{2}\beta(2) = 1 + \frac{1}{2}\left(23/8 + 9/4\right) = 57/16$

$E[\text{time}] = 5 \times 57/16 = 285/16 \approx 17.8\text{s}.$