Autograded 1:

Problem 4:

1. Random Variables:

N: The number of questions a student "knows" $\sim B(10, 0.6)$

Z: The number of correct answers among the questions the student guesses $\sim B(10-N, 0.5)$

$Y = N + Z$: Total number of correct answers

Threshold T : a student passes if $Y \geq T$

Conditional Probability: (Bayes's Rule)

* We want to compute $P(N<5 \mid Y \geq T) = \dfrac{P(N<5 \cap Y \geq T)}{P(Y \geq T)}$ → Numerator / → Denominator

Numerator: $P(N<5 \cap Y \geq T)$ → we need to sum over all $N<5$

↳ $= \sum_{n=0}^{4} P(N=n) \cdot P(Y \geq T / N=n)$ → Prob of guessing $\geq T-n$ correct answers when $N=n$

↳ prob of knowing n answers

Denominator: $P(Y \geq T)$

↳ Total probability of passing: $P(Y \geq T) = \sum_{n=0}^{10} P(N=n) \cdot P(Y \geq T \mid N=n)$

↳ $\binom{10}{n} \cdot p^n (1-p)^{10-n}$, where $p = 0{,}6$

Given $N=n$, $Y \geq T$ requires guessing $\geq T-n$ correctly from $10-n$ questions: $P(Y \geq T / N=n) = \sum_{k=\max(0, T-n)}^{10-n} \binom{10-n}{k} (0.5)^k \cdot (0{,}5)^{10-n-k}$

2. Find the smallest threshold T such that if $Y \geq T$ we are at least 90% certain that $N \geq 5$: $P(N \geq 5 \mid Y \geq T) \geq 0.90 \iff 1 - P(N<5 \mid Y \geq T) \geq 0.90 \iff P(N<5 \mid Y \geq T) \leq 0.1$

Problem 5:

1. Exponential ~~Distribution~~ Concentration : A RV Z satisfies exponential concentration if its deviation from the mean decay at an exponential rate :

$$P(Z - \mathbb{E}[Z] \geq \varepsilon) \leq C_1 \cdot e^{-C_2 n \cdot \varepsilon^2} \quad \text{or} \quad C_3 \cdot e^{-C_4 n(\varepsilon+1)}$$

* This is typically for sums / avg of sub-Gaussian or sub-Exponential RV that exhibit strong tail decay properties.

(2) Weaker Concentration: A RV Z satisfies weaker concentration if its deviation from the mean decay polynomially : $P(Z - \mathbb{E}[Z]) \geq \varepsilon) \leq \frac{C_1}{n \cdot \varepsilon^2}$

* This occurs for RV with finite variance but without strong tail decay properties like those of sub-Gaussian or sub-Exponential RV.

1) The empirical variance of iid RV with finite mean

↳ $S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ where X_i: RV. If only a finite mean is guaranteed (no sub-Gaussian/Exponential ⇒ the tails decay slowly ⇒ we get weaker polynomial concentration

2) The empirical variance of iid sub-Gaussian RV: A Sub Gaussian RV have strong tail decay. For X_i sub-Gaussian the empirical variance will inherit exponential concentration properties due to the exponential concentration of the terms $(X_i - E[X])^2$.

3) The empirical variance of iid sub-Exponential RV: Sub Exponential RV has strong tail decay though slightly weaker than sub-Gaussian. Empirical variance exhibits exp. concentr.

4) The empirical mean of iid sub-Gaussian RV:

↳ $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ of sub-Gaussian RV concentrates exponentially as sub-Gaussian RV exhibit exponential tail decay.

5) The empirical mean of iid sub-exponential RV: They also do ~~as~~ due to their strong tail decay properties.

6) The empirical mean of iid RV with finite variance:
If X_i only has a finite variance the Central Limit Theorem ensures that the e[illegible] mean concentrates but the concentration is weaker (polynomial / not exponential)

7) The empirical 3rd moment of iid RV with finite sixth moment:
Moments of higher order (3rd) depend on the higher moments of the distribution. If the sixth moment is finite, the 3rd moment will have weaker concentration as no sub-Gaussian / Exponential properties are assumed.

~~8)~~ The empirical 4th moment of iid sub-Gaussian RV. ∴ Sub-Gaussian RV have strong tail decay so higher moments (eg. 4th) will concentrate expontially.

9) The empirical mean of iid deterministic RV: If RV are deterministic there is no randomness and the empirical mean = actual mean (= perfect concentration).

10) The empirical 10th moment of iid Bernoulli RV: Bernoulli RV are sub Gaussian, thus even higher moments (eg. 10th) will concentrate exponentially

① 2, 3, 4, 5, ~~8~~, 9, 10

② ~~1~~, 2, 3, 4, 5, 6, 7, 8, 9, 10

* Correct explanation for 1 and 8

1) If only a finite mean is guaranteed (not even finite variance is specified) the behaviour of the empirical variance is poorly controlled bc the tails of the distribution can decay arbitrarily slowly. Without finite variance we cannot bound deviations in terms of $\frac{1}{n\varepsilon^2}$
↳ Does not concentrate

8) Sub-Gaussian RV exhibit strong tail decay. However the 4th moment is a non-linear function of the RV and does not inherit exponential concentration.
The 4th moment being derived from the squared terms, concentrates only in a weaker sense as its deviation are polynomially bounded.

Sub-Gaussian Distributions:

Bernoulli → X is bounded $Var(x) = p(1-p)$ & centered $(X-p)$: also sub-Gaussian

Uniform → tail bounds are tighter then Gaussian

Gaussian

Bounded RV (within an interval $[a,b]$)

Binomial (sum of independent sub-Gaussian vars - Bernoulli trials)

Sub-Exponential Distributions:

Exponential (λ)

Gamma (α, β)

Poisson (λ)

Chi-Squared $\sim X^2$

Log-Normal Distribution

* Sub-Gaussian ⊂ Sub-Exponential
↳ any sub-Gaussian RV is also sub-exponential (not vice-versa)

Assignment 2 - Problem 1

	To			
From	Downtown	Surburbs	Countryside	
Downtown	0.3	0.4	0.3	1
Surburbs	0.2	0.5	0.3	1
Countryside	0.4	0.3	0.3	1

$$P = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{bmatrix}$$

1. $P(\text{Surburbs} \to \text{Downtown in 2 steps}) = P \cdot P = P^2$

$$\begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{bmatrix}\begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{bmatrix} = \begin{bmatrix} 0.29 & 0.41 & 0.30 \\ \boxed{0.28} & 0.42 & 0.30 \\ 0.30 & 0.40 & 0.30 \end{bmatrix}$$

$P(\text{Surburbs} \to \text{Downtown in 2 steps}) = 0.28 = 28\%$

2. $P(\text{surburbs} \to \text{Downtown (first time) in 2 steps}) = P(S \to \underset{\text{except D}}{\text{Anywhere}} \to D)$

$= P(S \to S \to D) + P(S \to C \to D)$

$= P(S \to S) \cdot P(S \to D) + P(S \to C) \cdot P(C \to D) = 0.5 \cdot 0,2 + 0.3 \cdot 0,4 = 0,10 + 0.12 = 0,22 = 22\%$

3. Is it irreducible? If it possible to transition between any pair of states in a finite number of steps. For every pair of state i and j there exists a positive integer n such that the $(i;j)$-th entry of n-step ~~trat~~ transition matrix P^n is greater than zero. $P^n(i,j) > 0$

In the P (transition matrix) every entry has a non-zero prob of transitioning from one state (row) to another (column) in one step. It is a irreducible Markov's Chain

---

Non-irreducible MC:

eg. $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.3 & 0.7 \end{bmatrix}$ From S_1 is impossible to transition to S_2 or $S_3 \Rightarrow S_1$: absorbing state

States S_2 and S_3 are connected to each other. However is impossible to transition from S_2 or S_3 back to S_1 or from S_1 to S_2 or S_3

---

4. Stationary distribution $\pi = \{\pi_1, \pi_2, \pi_3\}$

$\pi_1 + \pi_2 + \pi_3 = 1 \quad \Leftrightarrow$

$0.3\pi_1 + 0.2\pi_2 + 0.4\pi_3 = \pi_1 \Leftrightarrow -0.7\pi_1 + 0.2\pi_2 + 0.4\pi_3 = 0$

$0.4\pi_1 + 0.5\pi_2 + 0.3\pi_3 = \pi_2 \Leftrightarrow -0.5\pi_2 + 0.4\pi_1 + 0.3\pi_3 = 0$

$0.3\pi_1 + 0.3\pi_2 + 0.3\pi_3 = \pi_3 \Leftrightarrow -0.7\pi_3 + 0.3\pi_2 + 0.3\pi_3 = 0$

5. Expected number of steps until first reaching downtown when starting in the suburbs
↳ $h_{S \to D}$ (expected hitting times)

$h_i = 1 + \sum_j P(i \to j) h_j$ if i = Downtown then $h_{Downtown} = 0$ bc expected hitting time to downtown from downtown is zero.

From the downtown state: $h_{Downtown} = 0$

From the suburbs: $h_{surburbs} = 1 + P(S \to D) \cdot h_{Downtown} + P(S \to S) h_{surburbs} + P(S \to C) \cdot h_{countryside}$

$= 1 + 0.2 \cdot 0 + 0.5\, h_{surburbs} + 0.3 \cdot h_{countryside} \Leftrightarrow 0.5 h_S - 0.3 h_C = -1 \quad (1)$

om the countryside state:

$h_{countryside} = 1 + P(C \to D) \cdot h_D + P(C \to S') \cdot h_{S'} + P(C \to C) \cdot h_C = 1 + 0.4 \cdot 0 + 0.3 \cdot h_{S'} + 0.$

$\Leftrightarrow \quad -0.3 h_{S'} + 0.7 h_C = -1 \quad (2)$

$(1)(2) \Rightarrow h_{surburbs} = 3.85$

Introdution to Data Science - Assignment 3:

$X \sim$ Uniform (B_1) $\quad Y = ||X||_2$ (Euclidean Norm)

) Find the distribution function of Y

The $X \sim$ Uniform (B_1) is a ~~ball~~ unit ~~ball~~ sphere of radius 1. $\{x \in B_1 : ||x|| = 1\}$
The density of the uniform random vector will be proportional to 1 over to the whole area of a ball ("circle"), that is π: $f_X(x) = \frac{1}{\pi}$ for $x \in B_1$

The norm $Y = ||X||_2$ is the distance between X and origin

~~The CDF of Y: $F_Y(y)$~~

Given we have a radius of y then the area of a disk is $\pi \cdot y^2$ and since we normalised the total probability to 1 using the norm we can say that the CDF of Y is: $F_Y(y) = P(Y \leq y) = \frac{\pi \cdot y^2}{\pi} = y^2$, for $0 \leq y \leq 1$.

To get the ~~first~~ PDF $f_Y(y)$ we need to ~~~~ differentiate the CDF ~~with~~ in terms of y

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(y^2) = 2y$$

b) find the distribution of $\ln(1/Y)$?

Let $W = \ln(1/Y) = \ln(Y^{-1}) = -\ln(Y)$

$$e^W = e^{-\ln(Y)} \Leftrightarrow e^W = \frac{1}{Y} \Leftrightarrow Y = e^{-W}$$

The PDF of W is given by:

$$f_W(w) = f_Y(y) \left|\frac{dy}{dw}\right|, \text{ where } y = e^{-w}$$

$\frac{dy}{dw} = -e^{-w}$ (1) and if we use $y = e^{-w}$ and the final equation on a) $\Rightarrow f_Y(y)$ becomes $f_Y(e^{-w}) = 2e^{-w}$ (2)

$$f_Y(y)\left|\frac{dy}{dw}\right| = 2e^{-w} \cdot e^{-w} = 2e^{-2w}, \quad w \geq 0.$$

therefore the PDF for $\ln(1/Y)$ is

$$f_W(w) = 2e^{-2w}, \quad w \geq 0.$$

c) Calculate $\mathbb{E}[\ln(1/Y)]$ first by using the distribution function of Y and then by the $\ln(1/Y)$

Distribution of Y:

we know $f_Y(y) = 2y$ for $0 \le y \le 1$

$E[\ln(1/Y)] = \int_0^1 \ln\left(\frac{1}{y}\right) \cdot f_Y(y) \cdot dy = \int_0^1 \ln\left(\frac{1}{y}\right) \cdot 2y\, dy = -2\int_0^1 y \cdot \ln(y) dy$ (1)

By using $u \cdot dv = uv - \int v \cdot du$

let $u = \ln(y)$ and $dv = y \cdot dy$. Then $du = \frac{1}{y}$ and $v = \frac{y^2}{2}$

$= \int y \cdot \ln(y) dy = \left[\frac{y^2}{2} \cdot \ln(y)\right]_0^1 - \int_0^1 \frac{y^2}{2} \cdot \frac{1}{y} dy = 0 - \left[\frac{y^2}{4}\right]_0^1 = \frac{-1}{4}$

~~Therefore~~ Hence$^{(1)}$: $E[\ln(1/Y)] = -2\left(-\frac{1}{4}\right) = \frac{1}{2}$

using distribution of $\ln(1/y)$:

From part b) we know $W = \ln(1/Y) \sim \text{Exponential}(\lambda = 2)$. The mean of such distribution is given by: $\frac{1}{\lambda}$

Hence $E[W] = \frac{1}{2} \Rightarrow E[\ln(1/y)] = \frac{1}{2}$.

4 d i) $v_1, v_2, \dots u_r$: From b) we showed that the matrix has a rank of r

↳ therefore we have r independent vectors

The Singular Value Decomposition is given by the formula: $A = U\Sigma V^T$, where...

~~To Perform SVD we need to compute U^TU to find the eigenvalues & eigenvectors~~

$U^TU = \left(\sum_{i=1}^r u_i u_i^T\right)^T \cdot \left(\sum_{j=1}^r u_i u_i^T\right) = \sum_{i=1}^r u_i u_i^T u_i u_i^T = \sum_{i=1}^r (u_i^T u_i)(u_i u_i^T) =$

The right singular vectors are the eigenvectors of the matrix u^Tu and left singular vectors are the eigenvectors of uu^T → $[U]$ → $[V]$

To find the right singular vectors of u, we need to compute U^TU which will give us the eigenvectors and eigenvalues.

~~$u_i u_i^T$ is a rank 1 matrix~~ Since u_i is a unit vector $(u_i^Tu_i = 1)$ we can simplify the $u^Tu = \sum^r u_i u_i^T$.

The matrix U^TU is symmetric and rank r which means that we have r eigenvalues and the eigenvectors of U^TU are the same as the eigenvectors of U (because $U = U^TU$)

The eigenvalues of U correspond to the r linearly independent vectors $u_1, u_2, \dots u_r$ as we know from $u_i u_i^T$ is the sum of each rank matrix

since $u_i^Tu_i = 1$ for each i this means $U^TU = U$

The matrix U^TU has a rank of r and its eigenvectors

$Y_1, Y_2, Y_3 \ldots Y_n$: be a sequence of IID discrete RVs where $P(Y_i=0) = 0.1$, $P(Y_i=1) = 0.3$, $P(Y_i=2) = 0.2$, $P(Y_i=3) = 0.4$. Let $X_n = \max\{Y_i, \ldots Y_n\}$. Let $X_0 = 0$ and verify that $X_0, X_1, \ldots X_n$ is a Markov chain. Find the transition matrix P.

Markov's Chain should satisfy the following property:

$$P(X_{t+1} = s \mid X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) = P(X_{t+1} = s \mid X_t = s_t)$$

The probability of X_{t+1} should depend to the propability of X_t and not ~~as~~ previous ones $(X_0, X_1, \ldots X_{t-1})$

Since ~~X_{n+1}~~ ~~can either~~ ~~X_{n+1} max~~ $X_n = \max\{Y_1, \ldots Y_n\}$ this means that it can X_{n+1} either ~~remain~~ remain as X_n or increase ~~to X_{n+1}~~

two possibilities 1) ~~if~~ If $X_n = a$, ~~..~~ then $X_{n+1} = a$, if $Y_{n+1} \le a$

2) if $X_{n+1} =$ ~~a~~ b if ~~X_{n+1}~~ $Y_{n+1} = b > a$

The prob of transitioning from X_n to X_{n+1} depends on the distribution of Y_i and the previous state X_n, $\Rightarrow$ Markov's ~~sta~~ chain satisfied.

b) States of $X_n = \{0, 1, 2, 3\}$ as it depends on ~~Y_i~~ Y_i

Let i and j be transition states (from i to j) where $i, j \in (0, 1, 2, 3)$

1. Transitioning (Let $i=0$) from $X_n = 0$ to:
   - $\rightarrow X_{n+1} = 0$ if $Y_{n+1} = 0$ prob $= 0.1$
   - $\rightarrow X_{n+1} = 1$ if $Y_{n+1} = 1$ prob $= 0.3$
   - $\rightarrow X_{n+1} = 2$ if $Y_{n+1} = 2$ prob $= 0.2$
   - $\rightarrow X_{n+1} = 3$ if $Y_{n+1} = 3$ prob $= 0.4$

First row $\{0,j\} = [0.1, 0.3, 0.2, 0.4]$

Transitioning from $X_n = 1$ to:
- $\rightarrow X_{n+1} = 1$ if $Y_{n+1} = 1$

For X_{n+1} to stay at 1 must be less than or equal to 1. This way Y_{n+1} doesn't introduce a new maximum greater than 1:

$X_{n+1} = 1$ if $Y_{n+1} = 0$ or $Y_{n+1} = 1$: prob $0.1 + 0.3 = 0.4$

$X_{n+1} = 2$ if $Y_{n+1} = 2$: prob 0.2

$X_{n+1} = 3$ if —"— $= 3$: prob. 0.4

Second row: $[0, 0.4, 0.2, 0.4]$

Transitioning from $X_n = 2$ to:
- $\rightarrow X_{n+1} = 2$: if $Y_{n+1} = 0$, $Y_{n+1} = 1$ or $Y_{n+1} = 2$ $= 0.1 + 0.3 + 0.2 = 0.6$
- $\rightarrow X_{n+1} = 3$ if $Y_{n+1} = 3$ prob. $= 0.4$

Third row $= [0, 0, 0.6, 0.4]$

Transition from $X_n = 3$ to:
- $\rightarrow$ if $X_n = 3 \Rightarrow$ then X_{n+1} will alway remain 3 since that is the max.

Last row $= [0, 0, 0, 1]$

$$\begin{array}{cc} & t-1 \\ P = & \begin{bmatrix} 0.1 & 0.3 & 0.2 & 0.4 \\ 0 & 0.4 & 0.2 & 0.4 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ t & \end{array}$$

state t.

state t-1

$P(Y_i=0)=0,1$

$P(Y_i=1)=0.3$

$P(Y_i=2)=0.2$

$P(Y_i=3)=0.4$

when $\underline{X_i=0}$: Y either 0 or 1.

From

$= 0,1+0.3=0.4$

From:

To

$Y=2$: $P(Y_i=2)=0.2$

$Y=3$; $P(Y_i=3)=0.4$

To

when $X_i=1$: (From) $Y=1$ (To) $\Rightarrow$ prob $=0,3$

$Y=2 \Rightarrow$ prob $=0.2$

$Y=3 \Rightarrow$ prob $=0.4$

when $X_i=2$: $Y=2 \Rightarrow$ prob $=0.2$.

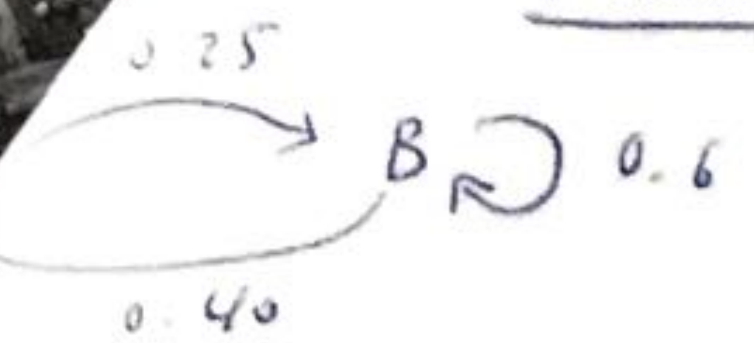
$Y=3 \Rightarrow$ prob $=0.4$

when $X_i=3$: $Y=3 \rightarrow$ only option as it can remain the same or increase but since $X=\max(Y_i)$ it will remain the same.

To

From:	0	1	2	3
0	0.4	0.4	0.2	0.4
1	0	0.3	0.2	0.4
2	0	0	0.2	0.4
3	0	0	0	1

Markov's Chains



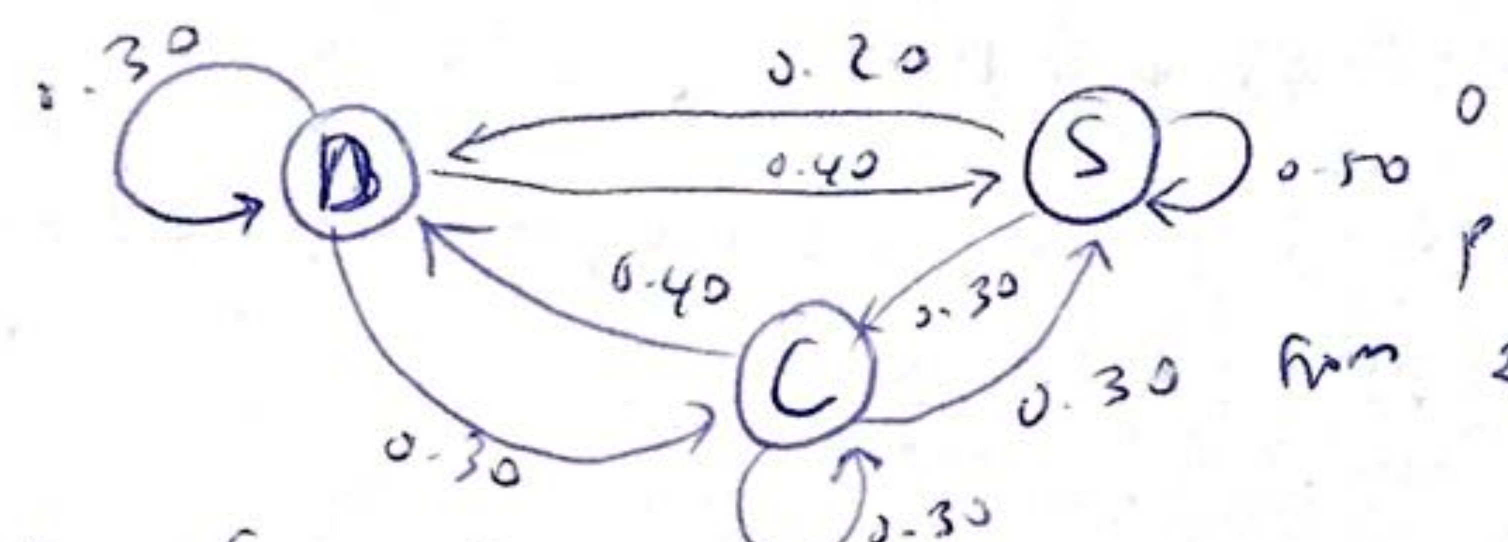
Transition Matrix = $\begin{pmatrix} 0.75 & 0.40 \\ 0.25 & 0.60 \end{pmatrix}$

$S_1 = P \cdot S_0$

$S_2 = P \cdot S_1 = P^2 \cdot S_0 \iff S_n = P^n \cdot S_0$

Ex. 1

	0 prob. transitioning to downtown	1 prob transitioning to surburbs	2 prob transitioning to countryside
0 Downtown	0.30	0.40	0.30
1 Surburbs	0.20	0.50	0.30
2 countryside	0.40	0.30	0.30



$$\begin{array}{cc} & \text{to} \\ & \begin{array}{ccc} 0\ D & 1\ S & 2\ C \end{array} \\ \begin{array}{c} 0\ D \\ P = 1\ S \\ \text{from } 2\ C \end{array} & \begin{bmatrix} 0.30 & 0.40 & 0.30 \\ 0.20 & 0.50 & 0.30 \\ 0.40 & 0.30 & 0.30 \end{bmatrix} \end{array}$$

a) starting from Surburbs → downdown in 2 steps

$$P^2 = P \cdot P = \begin{bmatrix} 0.30 & 0.40 & 0.30 \\ 0.20 & 0.50 & 0.30 \\ 0.40 & 0.30 & 0.30 \end{bmatrix} \begin{bmatrix} 0.30 & 0.40 & 0.30 \\ 0.20 & 0.50 & 0.30 \\ 0.40 & 0.30 & 0.30 \end{bmatrix} = \begin{bmatrix} 0.09+0.08+0.12 & 0.12+0.20+0.09 & 0.09+0.12+0.09 \\ 0.06+0.1+0.12 & 0.08+0.25+0.09 & 0.06+0.15+0.09 \\ 0.12+0.06+0.12 & 0.16+0.15+0.09 & 0.12+0.09+0.09 \end{bmatrix}$$

$$= \begin{array}{cc} & \begin{array}{ccc} D & S & C \end{array} \\ \begin{array}{c} D \\ S \\ C \end{array} & \begin{bmatrix} 0.29 & 0.41 & 0.30 \\ 0.28 & 0.42 & 0.30 \\ 0.30 & 0.40 & 0.30 \end{bmatrix} \end{array}$$

Probability = 0.28

b) starting in surburbs ⇒ downdown for the first time after 2 steps
there are two possible solutions:
1) $S \to S \to D$ prob. from $S \to S$ · prob from $S \to D$.
2) $S \to C \to D$. prob from $S \to C$ · prob from $C \to D$

1) $0.50 \cdot 0.20 = 0.10$
2) $0.30 \cdot 0.40 = 0.12$
} Prob. = 0.22.

c) Irreducible: when we can reach any state from any other state in a finite number of steps. The prob. should be > 0.
In this case it is irreducible

d) stationary distribution: $\pi = [\pi_D, \pi_S, \pi_C]$ such that $\pi \cdot P = \pi$

$\pi_C + \pi_D + \pi_S = 1$

$$\cancel{\pi = \begin{bmatrix} 0.30 & 0.40 & 0.30 \\ 0.20 & 0.50 & 0.30 \\ 0.40 & 0.30 & 0.30 \end{bmatrix} \begin{bmatrix} \pi_D \\ \pi_S \\ \pi_C \end{bmatrix}}$$

$0.30 \cdot \pi_D + 0.40 \cdot \pi_S + 0.30\, \pi_C = \pi_D$

0..

$$[\pi_D \quad \pi_S \quad \pi_C] \begin{bmatrix} 0.30 & 0.40 & 0.30 \\ 0.20 & 0.50 & 0.30 \\ 0.40 & 0.30 & 0.30 \end{bmatrix} = [\pi_D, \pi_S, \pi_C]$$

$0.30\pi_D + 0.40\pi_S + 0.30\pi_C = \pi_D$

$0.20\,\pi_D + 0.50\pi_S + 0.30\,\pi_C = \pi_S$

$0.40\pi_D + 0.30\pi_S + 0.30\pi_C = \pi_C$

$\pi_C + \pi_D + \pi_S = 1$

$\boxed{3.84}$

5) $E(S \to D) = \sum_{n=1}^{\infty} n \cdot$ Trans. Matrix

first calculate $E(S)$ $E(D) = 0$ (since you are here) $E(C)$

↳ start from surf. → D

$E(S) = 0.20 \cdot 0 + 0.5 \cdot (1 + E(S')) + 0.30 \cdot (1 + E(C)) \Leftrightarrow 0.50 + 0.50\,E(S) + 0.3 + 0.3\,E(C) = E(S')$

$\Leftrightarrow 0.5\,E(S') = 0.80 + 0.3\,E(C) \Leftrightarrow E(S) = 1.6 + 0.6\,E(C)$. (1)

$E(C) = 0.4 \cdot 0 + 0.3(1 + E(S)) + 0.30(1 + E(C)) \Leftrightarrow 0.30 + 0.30\,E(S') + 0.3 + 0.3\,E(C) = E(C)$

$\Leftrightarrow 0.7\,E(C) = 0.30\,E(S') + 0.6 \Leftrightarrow E(C) = 0.43\,E(S') + 0.86$ (2) 0.74

(1) + (2) $\Rightarrow E(C) = 0.43 \cdot [1.6 + 0.6\,E(C)] + 0.86 \Leftrightarrow E(C) = 0.69 + 0.26\,E(C) + 0.86$

$\Leftrightarrow E(C) = 2.09$.

(1) $E(S) = 1.6 + 0.6 \cdot 2.09 = 2.854$

$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right)$. $\sum_{i=1}^{n} \log(f(x_i; \mu, \sigma))$.

$f(x, \lambda) = \frac{1}{24} \cdot \lambda^5 \cdot x^4 \cdot \exp(-\lambda x)$.

$L(\lambda) = \prod_{i=1}^{n} f(x_i; \lambda) = \prod_{i=1}^{n} \frac{1}{24} \lambda^5 \cdot x^4 \exp(-\lambda x)$

$\log(L(\lambda)) = \sum_{i=1}^{n} \log\left(\cancel{\frac{1}{24} \lambda^5 x_i^4 \exp(-\lambda x_i)}\right) \left(\frac{1}{24}\lambda^5 \cdot x_i^4 \cdot \exp(-\lambda \cdot x_i)\right)$

$\log L(\lambda) = \sum_{i=1}^{n}\left(\log\frac{1}{24} + 5\log\lambda + 4\log x_i - \lambda x_i\right) = n\log\frac{1}{24} + 5n \cdot \log\lambda + 4\sum_{i=1}^{n}\log x_i - \lambda\sum_{i=1}^{n} x_i$

$\frac{d \log L(\lambda)}{d\lambda} = \frac{5n}{\lambda} - \sum_{i=1}^{n} x_i = 0$

$\sum_{i=1}^{n} x_i = \frac{5n}{\lambda} \Leftrightarrow \hat{\lambda} = \frac{5n}{\sum_{i=1}^{n} x_i}$ ⇝ MLE for λ.