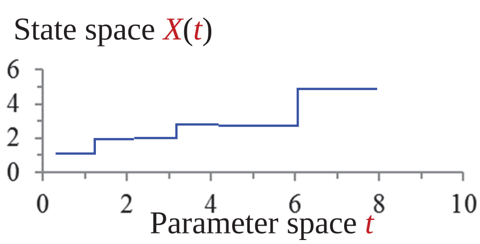
Session 9. Markov Chain Models

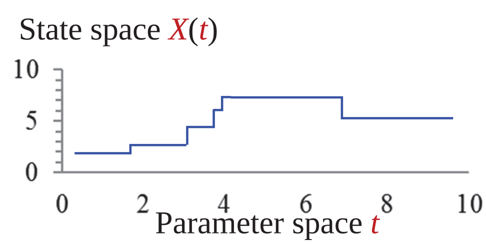
* Stochastic Process

- $X(t)$, $t \in T$ is a collection of random variable
  where t is often interpreted as *time*,
  $X(t)$ is the *state* of the process at time t,
  T is called the *index set* of the stochastic process.
- Parameter space:
  The set of possible values of the indexing parameter.
- State space:
  The set of all possible values that $X(t)$ can assume.
- Classification:

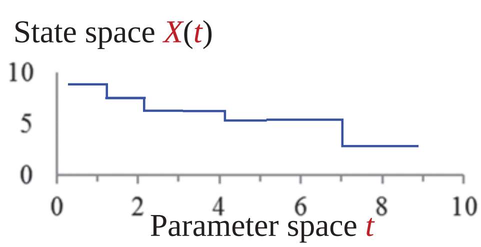
Various Cases		Parameter Space	
		Discrete	Continuous
State Space	Discrete	**Case I**	Case II
	Continuous	Case III	Case IV



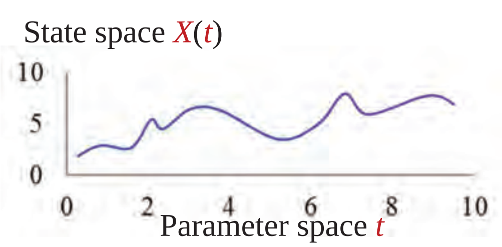
Case I



Case II



Case III



Case IV

A Transition Probabilities

* Definition

- A *discrete-time* stochastic process is a Markov Chain if

  $$P[X_{t+1}=i_{t+1} \mid X_t=i_t, X_{t-1}=i_{t-1},\ldots, X_0=i_0] = P[X_{t+1}=i_{t+1} \mid X_t=i_t]$$

  for t = 0, 1, 2,..., and all states.

- *First-order* dependence:

  The probability distribution of the state at time t+1 depends only on the state at time t.

- *Transition* probability: $p_{ij} = P[\, X_{t+1}=j \mid X_t=i \,]$

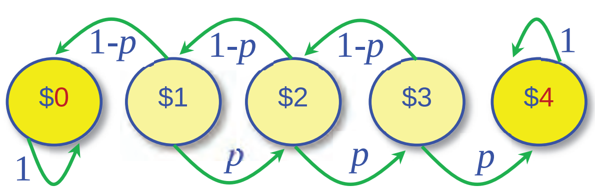
  The probability that the state will be changed from i to j.

- Probability distribution of the initial states: $\mathbf{q} = [\, q_1, q_2,\ldots,q_s \,]$

  The probability that the Markov chain is in state i at time 0.

Ex 1] The gambler's ruin: At time 0, I have \$2. At time 1, 2, ..., I play a game in which I bet \$1. With probability p, I win the game, and with probability 1-p, I lose the game. My goal is to increase my capital to \$4, and as soon as I do, the game is over. The game is also over if my capital is reduced to \$0.

- Transition diagram



- Transition probability

	0	\$1	\$2	\$3	\$4
0	1				
\$1	1-p		p		
\$2		1-p		p	
\$3			1-p		p
\$4					1

- Initial probabilities: $\mathbf{q} = [\, 0, 0, 1, 0, 0 \,]$
- Stage and state? Recurrent or absorbing states?

Ex 2] Maintenance problem: Laundromat has two washing machines. During any day, each machine that is working at the beginning of the day has a 1/3 chance of breaking down. If a machine breaks down during the day, it is sent to a repair facility and will be working two days after it breaks down. (e.g., If a machine breaks down during day 3, it will be working at the *beginning* of day 5.)

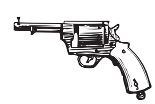
- State: Number of washing machines in working condition
- Transition matrix

P =

	0	1	2
0			1
1		1/3	2/3
2	1/9	4/9	4/9

Ex 3] Russian roulette: The game is played with a 6-shooter revolver with one blank cartridge. You spin the cylinder once. You then pull the trigger as many times as you want. Whenever you survive, you get paid $1. When do you want to stop and abandon the game?

- State: You survived the *i*th pull of the trigger.



- Transition probabilities p_{ij} and rewards r_i

p_{ij}	0	1	2	3	4	5	X	Reward r_i
0		5/6					1/6	0
1			4/5				1/5	$1
2				3/4			1/4	$2
3					2/3		1/3	$3
4						1/2	1/2	$4
5							1	$5
X							1	0

B. n-Step Transition Probabilities

* Conditional Probability

- If a Markov chain is in state i at time t, what is the probability that, n periods later, the Markov chain will be in state j?

$$p_{ij}(n) = \mathrm{P}[\, X_{t+n} = j \mid X_t = i \,] = \mathrm{P}[\, X_n = j \mid X_0 = i \,]$$

- Chapman – Kolmogorov equation:

$$p_{ij}(m+n) = \sum_{k=0}^{\infty} p_{ik}(m)\; p_{kj}(n)$$

for all n, $m \geq 0$ and all i and j.

- By induction, we can show that

$$\left\| p_{ij}(n) \right\| \;=\; \boldsymbol{P}^n.$$

- That is, the n-step transition matrix can be obtained by *multiplying* the matrix **P** by itself n times!

* Unconditional Probability

- The *initial* probability distribution, $\mathbf{q} = [\, q_1, q_2, \ldots, q_s \,]$, is the probability that the chain is in state i at time 0.

- Then, the *unconditional* probability that the state at time n is j is

$$p_{\bullet j}(n) = \sum_{i=1}^{s} q_i \;\, p_{ij}(n)$$

Ex 1] The Soda Example

Consider two types of soda: Soda A and Soda B. Given that a person last purchased soda A, there is a 90% chance that her next purchase will be soda A. Given that a person last purchased soda B, there is an 80% chance that her next purchase will be also soda B.

(a) Formulate a transition probability matrix.

$$\mathbf{P} = \begin{array}{c|cc|} & A & B \\ A & & \\ B & & \end{array}$$

(b) If a person is currently a soda A purchaser, what is the probability that she will purchase soda A two purchases from now?

$$\mathbf{P}^2 = \mathbf{PP} = \begin{array}{c|cc|} & A & B \\ A & & \\ B & & \end{array}$$

(c) If a person is currently a soda A purchaser, what is the probability that she will purchase soda A three purchases from now?

$$\mathbf{P}^3 = \mathbf{PP}^2 = \begin{array}{c|cc|} & A & B \\ A & & \\ B & & \end{array}$$

(d) Suppose 60% of all people now drink soda A, and 40% now drink soda B. Three purchases from now, what fraction of all purchasers will be drinking soda A?

$$\mathbf{qP}^3 =$$

Ex 2] Where to Live?

Each American family is classified as living in an urban, suburban, or rural location during a given year, 15% of all *urban* families move to a suburban location, and 5% moves to a rural location; also, 6% of all *suburban* families move to an urban location, and 4% move to a rural location; finally, 4% of all *rural* families move to an urban location, and 6% move to a suburban location.

(a) If a family now lives in an *urban* location, what is the probability that it will live in an *urban* area two years from now?

$$\mathbf{P} = \begin{array}{l|ccc|} & \text{Urban} & \text{Suburban} & \text{Rural} \\ \text{Urban} & & & \\ \text{Suburban} & & & \\ \text{Rural} & & & \end{array}$$

$$\mathbf{P}^2 = \begin{array}{l|ccc|} & \text{Urban} & \text{Suburban} & \text{Rural} \\ \text{Urban} & & & \\ \text{Suburban} & & & \\ \text{Rural} & & & \end{array}$$

(b) Suppose that at present, 40% of all families live in an urban area, 35% live in a suburban area, and 25% live in a rural area. Two years later from now, what percentage of American families will live in an *urban* area?

$\mathbf{q} =$

$\mathbf{q}\,\mathbf{P}^2 =$

(c) What problems might occur if this model were used to predict the future population distribution of the United States?

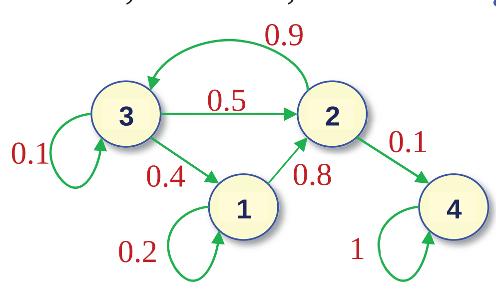
C. Steady-State Probabilities

* Classification of States in a Markov Chain

- A state j is *reachable* (accessible) from a state i if there is a path leading from i to j.
- Two states i and j are said to *communicate* if j is reachable from i and i is reachable from j.
- A set of states S in a Markov chain is a *closed set* (class) if no state outside of S is reachable from any state in S.
- A state i is an *absorbing state* if $p_{ii} = 1$.
- A state i is a *transient state* if there exists a state j that is reachable from i, but the state i is not reachable from state j.
- If a state is not transient, it is called a *recurrent state*.
- If all states in a chain are recurrent, aperiodic, and communicate with each other, the chain is said to be *ergodic*.

Ex] Determine whether the following Markov chain is ergodic. Also determine if the states are recurrent, transient, or absorbing.

$$\mathbf{P} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & .9 & .1 \\ 0.4 & 0.5 & 0.1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



* Steady-State Probabilities

- The *steady-state* probabilities of a irreducible, ergodic Markov chair are used to describe its long-run behavior.

- Let **P** be the transition matrix for an s-state ergodic chain. Then, there exists a vector $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_s]$ such that

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \pi_1 & \pi_2 & . & \pi_s \\ \pi_1 & \pi_2 & . & \pi_s \\ . & . & . & . \\ \pi_1 & \pi_2 & . & \pi_s \end{bmatrix}$$

$\pi_j = \sum_{k=1}^{s} \pi_k p_{kj}$, for $j = 1, 2, \ldots, s$

In matrix form, $\boldsymbol{\pi} = \boldsymbol{\pi}\,\mathbf{P}$

For the gambler's ruin problem, why is it unreasonable to talk about *steady-state* probabilities?

- Intuitive interpretation of steady-state probabilities:

$$\pi_j(1 - p_{jj}) = \sum_{k \neq j}^{s} \pi_k p_{kj}$$

Probability that a particular transition leaves state j
= Probability that a particular transition enters state j.

* Mean First Passage Time

m_{ij} = Expected number of transitions before we first reach state j, given that we are currently in state i.

$$m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj} \quad \text{and} \quad m_{ii} = \frac{1}{\pi_i}$$

* How to Find Steady-State Probabilities?

- The *steady-state* probability vector **π** of a Markov chain **P** can be found from the following equation:

$$\boldsymbol{\pi} = \boldsymbol{\pi}\,\mathbf{P}$$

where $\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdot & p_{1s} \\ p_{21} & p_{22} & \cdot & p_{2s} \\ \cdot & \cdot & \cdot & \cdot \\ p_{s1} & p_{s2} & \cdot & p_{ss} \end{bmatrix}$ is a $s \times s$ square matrix

and $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_s]$ is a *row* vector.

(a) Direct method: Gaussian elimination procedure

- The resulting set of equations is not linearly independent and one of the equations is redundant.
- To yield a unique, positive solution, a normalization condition, $\pi_1+\pi_2+\ldots+\pi_s = 1$, has to replace one of the equations.

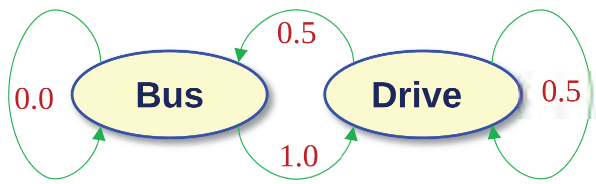
(b) Iterative method: Power method

- Start with an initial *row* vector $\mathbf{v}^{(0)}$.
- Repeatedly multiply it by the transition probability matrix **P** until convergence to **v** is reached.

$$\boldsymbol{v}^{(i)} = \boldsymbol{v}^{(i-1)}\boldsymbol{P} = \boldsymbol{v}^{(0)}\boldsymbol{P}^{i}$$

- To yield the final result of the steady-state probability vector **π**, only a *re-normalization* remains to be performed.

Ex 1] Bus example: A student either drives his car or catches a bus to school each day. Suppose he never goes by bus two days in a row; but if he drives to school, then the next day he is just as likely to drive again as he is to travel by bus.



	Bus	Drive
Bus	**0.0**	**1.0**
Drive	**0.5**	**0.5**

(a) Direct method (Use Excel-Solver!)

$\pi_1 = \pi_1 * 0.0 + \pi_2 * 0.5$

$\pi_2 = \pi_1 * 1.0 + \pi_2 * 0.5$ with

$\pi_1 + \pi_2 = 1$

Thus, $\pi_1 = 1/3$ and $\pi_2 = 2/3$

(b) Power method

$\mathbf{v}^{(0)} = [1, 1]$, which is an arbitrary initial vector.

$\mathbf{v}^{(1)} = [1, 1]\begin{bmatrix} 0 & 1 \\ .5 & .5 \end{bmatrix} = [.5,\ 1.5]$

$\mathbf{v}^{(2)} = [.5, 1.5]\begin{bmatrix} 0 & 1 \\ .5 & .5 \end{bmatrix} = [.75,\ 1.25]$

….

$\mathbf{v}^{(10)} = [.666, 1.334]\begin{bmatrix} 0 & 1 \\ .5 & .5 \end{bmatrix} = [.667,\ 1.333]$

After normalization, we finally have

$\pi_1 = .667/(.667+1.333) =$

$\pi_2 = 1.333/(.667+1.333) =$

(c) Mean first passage time with π = [1/3, 2/3]

$m_{11} = 1/\pi_1 =$

$m_{12} = 1 + p_{11}\, m_{12} =$

$m_{21} = 1 + p_{22}\, m_{21} =$

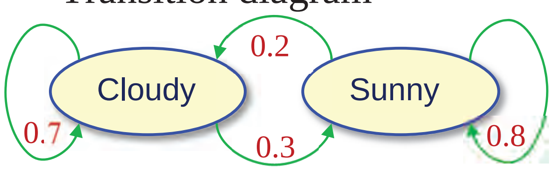
$m_{22} = 1/\pi_2 =$

$m_{11} = 3.$ If you take a bus today,
you will take a bus again three days later.

$m_{12} = 1.$ If you take a bus today,
you will drive tomorrow.

Ex 2] Local weather: Find the limiting probabilities and the mean first passage time.

- Transition diagram



- Transition matrix

	Cloudy	Sunny
Cloudy	0.7	0.3
Sunny	0.2	0.8

- Limiting probabilities

$\pi_1 =$

$\pi_2 =$ with $\pi_1 + \pi_2 = 1$

Thus, $\pi_1 =$ and $\pi_2 =$

- Mean first passage time

$m_{11} = 1/\pi_1 =$
$m_{12} = 1 + p_{11}\, m_{12} =$
$m_{21} = 1 + p_{22}\, m_{21} =$
$m_{22} = 1/\pi_2 =$

D. Markov Chain with Absorbing States

* Absorbing and Transient States

- If we begin in a *transient* state, then eventually we are sure to leave the transient state and end up in one of the *absorbing* states.

- *Canonical* representation: $\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$

- *Fundamental* Matrix: $\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1} = \sum_{r=0}^{\infty} \mathbf{Q}^r$

* Decision Problems

- **Question 1**. If the chain begins in a given transient state, and before we reach an absorbing state, what is the **expected number** of times that each state will be entered? (i.e., How many periods do we expect to spend in a given transient state before *absorption* takes place?)

Answer: If we are at present in transient state t_i, the expected number of periods that will be spent in transient state t_j before absorption is the ij^{th} element of the fundamental matrix, $\mathbf{M}=(\mathbf{I}-\mathbf{Q})^{-1}$.

- **Question 2**. If a chain begins in a given transient state, what is the **probability** that we end up in each absorbing state?

Answer: If we are at present in transient state t_i, the probability that we will eventually be absorbed in absorbing state a_j is the ij^{th} element of the matrix $(\mathbf{I}-\mathbf{Q})^{-1}\,\mathbf{R}$.

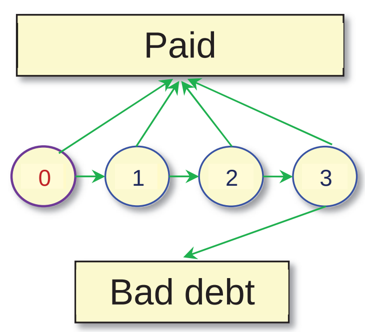
Ex 1] Accounts Receivable

The accounts receivable situation of a firm is often modeled as an absorbing Markov chain. Suppose a firm assumes that an account is uncollectable if the account is more than 3 months overdue. Then at the beginning of each month, each account may be classified into one of the following states:

t_1	New account
t_2	Payment on account is 1 month overdue.
t_3	Payment on account is 2 months overdue.
t_4	Payment on account is 3 months overdue.
a_5	Account has been paid.
a_6	Account is written off as bad debt.

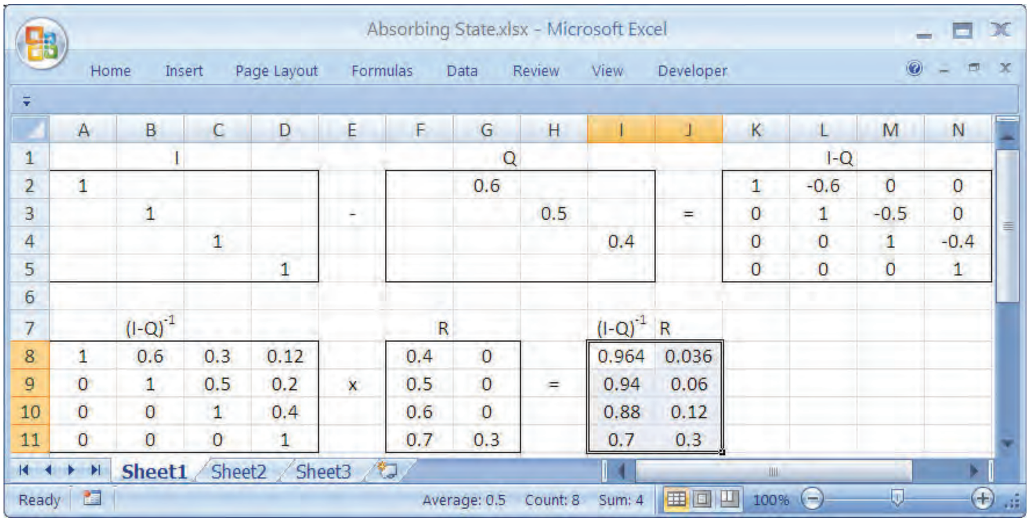
Suppose that past data indicate that the following Markov chain describes how the status of an account changes from one month to the next month.

$$\mathbf{P} = \begin{array}{r} New \\ 1\ month \\ 2\ month \\ 3\ month \\ Paid \\ Bad\ debt \end{array} \left[\begin{array}{cccc|cc} 0 & .6 & 0 & 0 & .4 & 0 \\ 0 & 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & 0 & .4 & .6 & 0 \\ 0 & 0 & 0 & 0 & .7 & .3 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$



$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \text{ where } \mathbf{Q} = \begin{bmatrix} 0 & .6 & 0 & 0 \\ 0 & 0 & .5 & 0 \\ 0 & 0 & 0 & .4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{R} = \begin{bmatrix} .4 & 0 \\ .5 & 0 \\ .6 & 0 \\ .7 & .3 \end{bmatrix}.$$

▪ Microsoft Excel



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		I					Q					I-Q		
2	1						0.6				1	-0.6	0	0
3		1			-			0.5		=	0	1	-0.5	0
4			1						0.4		0	0	1	-0.4
5				1							0	0	0	1
6														
7		$(I-Q)^{-1}$				R			$(I-Q)^{-1}$	R				
8	1	0.6	0.3	0.12		0.4	0		0.964	0.036				
9	0	1	0.5	0.2	x	0.5	0	=	0.94	0.06				
10	0	0	1	0.4		0.6	0		0.88	0.12				
11	0	0	0	1		0.7	0.3		0.7	0.3				

(a) If the firm's sales average is $100,000 per month, how much money per year will go uncollected?

(b) What is the probability that a one-month overdue account will eventually become a bad debt?

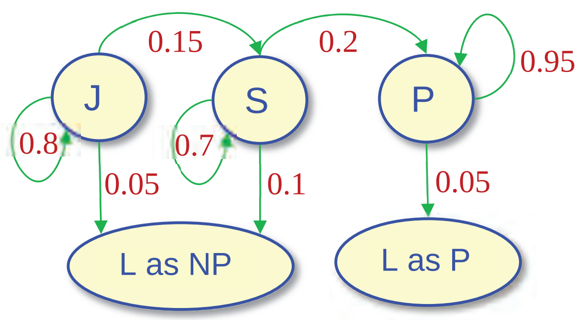
(c) What is the average time that a new account will eventually be collected or written off as bad debt.

Ex 2] Work-Force Planning

A law firm in Memphis employs three types of lawyers: Junior lawyers, senior lawyers, and partners.

During a given year, there is a 0.15 probability that a junior lawyer will be promoted and become a senior lawyer and there is a 0.05 probability that he or she will leave the firm. Also, there is a 0.20 probability that a senior lawyer will be promoted to partner and there is a 0.10 probability that he or she will leave the firm. There is also a 0.05 probability that a partner will leave the firm. The firm never demotes a lawyer.

	Junior	Senior	Partner	Leave as NP	Leave as P
Junior	0.80	0.15		0.05	
Senior		0.70	0.20	0.10	
Partner			0.95		0.05
Leave as NP				1	
Leave as P					1



$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

$$\text{where } \mathbf{Q} = \begin{bmatrix} 0.80 & 0.15 & 0 \\ 0 & 0.70 & 0.20 \\ 0 & 0 & 0.95 \end{bmatrix} \text{ and } \mathbf{R} = \begin{bmatrix} 0.05 & 0 \\ 0.10 & 0 \\ 0 & 0.05 \end{bmatrix}$$

	A	B	C	D	E	F	G	H	I	J	K
1		I				Q				I-Q	
2	1				0.8	0.15			0.2	-0.15	0
3		1		-		0.7	0.2	=	0	0.3	-0.2
4			1				0.95		0	0	0.05
5											
6		$(I-Q)^{-1}$			R			$(I-Q)^{-1}$ R			
7	5	2.5	10		0.05	0		0.5	0.5		
8	0	3.333	13.33	x	0.1	0	=	0.33333	0.66667		
9	0	0	20		0	0.05		0	1		

(a) What is the probability that a newly hired junior lawyer makes it to partner?

(b) What is the average length of time that a newly hired junior lawyer spends working for the firm?

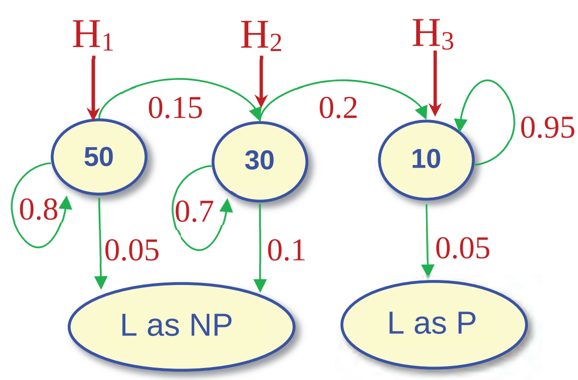
(c) What is the average length of time that a partner spends with the firm (as a partner)?

(d) Suppose the firm's long-term goal is to employ N_1=50 junior lawyers, N_2=30 senior lawyers, and N_3=10 partners. To achieve this steady-state census, how many lawyers of each type H_i should the firm hire each year?

- Number of people entering group i during each period $\quad H_i + \sum_{k \neq i} N_k p_{ki}$
- Number of people leaving group i during each period $\quad N_i \sum_{k \neq i} p_{ik}$
- Steady-state census $\quad H_i + \sum_{k \neq i} N_k p_{ki} = N_i \sum_{k \neq i} p_{ik}$

P =

	Junior	Senior	Partner	Leave
Junior	0.80	0.15		0.05
Senior		0.70	0.20	0.10
Partner			0.95	0.05



- Junior $\quad H_1 = (0.15+0.05) \times 50$
- Senior $\quad H_2 + 0.15 \times 50 = (0.2+0.1) \times 30$
- Partner $\quad H_3 + 0.2 \times 30 = 0.05 \times 10$

- Unique solution $\quad H_i =$

*E. Markov Decision Process**

I. Markov Process

- A state s_t is Markov if and only if

  $$P[s_{t+1} \mid s_t] = P[s_{+1} \mid s_t, s_{t-1}, s_{t-2}, \ldots, s_2, s_1]$$

- The state s_t is a *sufficient* statistic of the future;
  It captures all relevant information from the prior history.
  Once the state s_t is known, the history may be thrown away.

- A time-*homogeneous* Markov process is a sequence of random states s_1, s_2, … with the Markov property.

II. Markov Reward Process

- Each state has a reward (or cost), $\{r_1, r_2, \ldots, r_N\}$.

- At state s_i, you get given reward r_i and randomly move to another state s_j according to the transition probability p_{ij}.
  All future rewards are *discounted* by β where $0 < \beta < 1$.

- The state value function $v(s_i)$ is the expected *discounted* reward starting from state s_i.

  $$v(s_i) = r_i + \beta \sum_{j=1}^{N} v(s_j) p_{ij}, \qquad \text{(Bellman equation)}$$

  which is $\mathbf{v} = \mathbf{r} + \beta\, \mathbf{P}\, \mathbf{v}$ in matrix form

- We can easily get the *closed form* expression for $\mathbf{v}^*$ with matrix inversion:

  $$\mathbf{v}^* = (\mathbf{I} - \beta\mathbf{P})^{-1}\, \mathbf{r}$$

- Alternatively, we can find the answer by value iterations:

  $\mathbf{v}^{[t]} = \mathbf{r} + \beta\mathbf{P}\, \mathbf{v}^{[t-1]}$, with the initial values $\mathbf{v}^{[0]} = \mathbf{r}$.

Ex 1] Machine Replacement

At the beginning of each week, a machine is in one of four conditions, s_i = {excellent, good, average, or poor}. The quality of a machine deteriorates over time. The transition probabilities p_{ij} along with the weekly revenue r_i earned by a machine in each type of condition are given below:

p_{ij}	s_1	s_2	s_3	s_4	Revenue r_i
s_1: Excellent	0.7	0.3	-	-	\$100
s_2: Good	-	0.7	0.3	-	\$80
s_3: Average	-	-	0.6	0.4	\$50
s_4: Poor	-	-	-	1.0	\$10

Find the expected discounted rewards if the discount rate is β=0.9.

- **Method 1**: Matrix inversion

$$\mathbf{v}^* = (\mathbf{I} - \beta\mathbf{P})^{-1}\,\mathbf{r} = \begin{bmatrix} 527.61 \\ 352.64 \\ 186.96 \\ 100.00 \end{bmatrix}$$

If we begin with an *excellent* machine (s_1), the expected discounted reward of \$527.61 could be earned.

- **Method 2**: Value iteration

$$\mathbf{v}^{[t]} = \mathbf{r} + \beta\mathbf{P}\,\mathbf{v}^{[t-1]}$$

Iteration, i	0	1	2	3	..	81	82
$v(s_1)^{[i]}$	100	184.60	255.15	312.70	..	527.59	527.59
$v(s_2)^{[i]}$	80	143.90	192.42	228.32	..	352.63	352.63
$v(s_3)^{[i]}$	50	80.60	100.36	113.95	..	186.94	186.94
$v(s_4)^{[i]}$	10	19.00	27.10	34.39	..	99.98	99.98

III. Markov Decision Process

- At the current state s_i, choose one of the available *actions*, $a_k \in \{a_1, a_2, \ldots, a_M\}$, and receive the *reward* $r_i(k)$.

- If you choose action a_k when the state is s_i, you'll randomly move to the next state s_j with probability $p_{ij}(k)$.

- All *future rewards* are discounted by β per period. When the state is s_i, how can you find the optimal policy that maximizes the expected *discounted* total rewards?

▪ **Method 1**: Value iteration

- Use the iterative method to compute the value $v^*(s_i)$ for all i:

$$v^{[t]}(s_i) = \max_k \left[r_i(k) + \beta \sum_{j=1}^{N} v^{[t-1]}(s_j)\, p_{ij}(k) \right], \text{ for all } i,$$

with the initial values, $v^{[0]}(s_i) = r_i(k)$. Let $v^*(s_i) = v^{[\infty]}(s_i)$.

- When we are in state s_i, the best action a_k is the one that maximizes

$$r_i(k) + \beta \sum_{j=1}^{N} v^*(s_j)\, p_{ij}(k), \text{ for all } k.$$

▪ **Method 2**: Linear programming

- The optimal policy for a problem of *maximizing* the discounted total rewards can be found by solving the following LP (all variables v_j are urs):

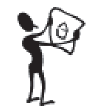
$$\text{Min } z = \sum_{j=1}^{N} v_j$$

$$\text{s.t. } v_i - \beta \sum_{j=1}^{N} v_j\, p_{ij}(k) \geq r_i(k) \text{ for all } i.$$

- If the shadow price of a constraint for action a_k and state s_i has a *non-zero* value, then action a_k is optimal in state s_i.

Ex 2] Machine Replacement (*revisited*)

After observing the condition of the machine at the beginning of the week, we have the option of *instantaneously* replacing it with an excellent machine, which costs $200. (In such a case, the revenue with a newly replaced machine is r_1 - $200 = -$100, no matter what type of machine we had at the beginning of the week.) Find the optimal replacement policy.



- **Method 1**: Value iteration

(i) If you take a_1 (Not replace)

$p_{ij}(1)$	s_1	s_2	s_3	s_4	$r_i(1)$
s_1	0.7	0.3	-	-	\$100
s_2	-	0.7	0.3	-	\$80
s_3	-	-	0.6	0.4	\$50
s_4	-	-	-	1.0	\$10

(ii) If you take a_2 (Replace)

$p_{ij}(2)$	s_1	s_2	s_3	s_4	$r_i(2)$
s_1	0.7	0.3	-	-	-\$100
s_2	0.7	0.3	-	-	-\$100
s_3	0.7	0.3	-	-	-\$100
s_4	0.7	0.3	-	-	-\$100

- Results after 120 value iterations:

	a_1 (Not replace)	a_2 (Replace)	Max	Optimal action
$v^*(s_1)$	690.23	490.23	690.23	a_1
$v^*(s_2)$	575.50	490.23	575.50	a_1
$v^*(s_3)$	492.36	490.23	492.36	a_1
$v^*(s_4)$	451.21	490.23	490.23	a_2

- Optimal replacement policy:

Replace (a_2) if the machine is in *poor* condition (s_4).

If the revenue r_3 has been decreased from $50 to $30?

	a_1 (Not replace)	a_2 (Replace)	Max	Optimal action
$v^*(s_1)$	687.81	487.81	687.81	a_1
$v^*(s_2)$	572.19	487.81	572.19	a_1
$v^*(s_3)$	469.03	487.81	487.81	a_2
$v^*(s_4)$	449.03	487.81	487.81	a_2

- **Method 2**: Linear programming

Min $z = v_1 + v_2 + v_3 + v_4$

(1) $v_1 - 0.9\,(0.7v_1 + 0.3v_2) \geq 100$ (Not replace when s_1)
(2) $v_2 - 0.9\,(0.7v_2 + 0.3v_3) \geq 80$ (Not replace when s_2)
(3) $v_3 - 0.9\,(0.6v_3 + 0.4v_4) \geq 50$ (Not replace when s_3)
(4) $v_4 - 0.9\,(1v_4) \geq 10$ (Not replace when s_4)

(5) $v_1 - 0.9\,(0.7v_1 + 0.3v_2) \geq -100$ (Replace when s_1)
(6) $v_2 - 0.9\,(0.7v_1 + 0.3v_2) \geq -100$ (Replace when s_2)
(7) $v_3 - 0.9\,(0.7v_1 + 0.3v_2) \geq -100$ (Replace when s_3)
(8) $v_4 - 0.9\,(0.7v_1 + 0.3v_2) \geq -100$ (Replace when s_4)

and all variables v_j are *unrestricted in sign*.

- LP solution from Microsoft Excel - Solver

$\mathbf{v}^* = [690.23, 575.50, 492.36, 490.23]$

which agree with those found via the method of value iteration.

- Optimal policy

a_1: Not replace
Shadow prices of constraints (1), (2), (3) are *non-zero*.
Shadow price of constraint (4) is zero.

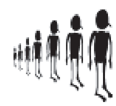
a_2: Replace
Shadow prices of constraints (5), (6), and (7) are *zero*.
Shadow price of constraint (8) is non-zero.

Thus, the optimal policy is to replace the machine (a_2) if and only if the machine is in *poor* condition (s_4). If we begin with an excellent machine (s_1). the expected discounted reward of \$690.23 could be earned.

Ex 3] Secretary Problem with n=4 (*revisited*)

* Formulation

- Stage i = The ith interview, where i=1, 2, 3, 4, and fail.
- State s_i = The ith choice is a candidate.
- Action a_k = { a_1: Continue, a_2: Stop }



- Transition probabilities:

  $p_{ij}(a_1)=\frac{i}{j-1}\frac{1}{j}$ for $1\le i<j\le n$. $p_{ij}(a_2) = 1$ for $i=j$.

- Rewards:

  $r_i(a_1) = 0$ $r_i(a_2) = \frac{i}{n}$ for i=1,2,..,n.

* Method 1: Value iteration

(i) If you take a_1 (Continue)

$p_{ij}(1)$	s_1	s_2	s_3	s_4	Fail	$r_i(1)$
s_1		1/2	1/6	1/12	1/4	0
s_2			1/3	1/6	2/4	0
s_3				1/4	3/4	0
s_4					1	0
Fail					1	0

(ii) If you take a_2 (Stop)

$p_{ij}(2)$	s_1	s_2	s_3	s_4	Fail	$r_i(2)$
s_1	1					0.25
s_2		1				0.50
s_3			1			0.75
s_4				1		1.00
Fail					1	0.00

- Results after 3 value iterations:

	a_1 (Continue)	a_2 (Stop)	Max	Optimal action
$v^*(s_1)$	0.4583	0.4583	0.4583	a_1
$v^*(s_2)$	0.4167	0.5000	0.5000	a_2
$v^*(s_3)$	0.2500	0.7500	0.7500	a_2
$v^*(s_4)$	0.0000	1.0000	1.0000	a_2

- Optimal policy:

  Learning set = { s_1 } and *action* set = {s_2, s_3, s_4}.

  P[Win | n=4] = 0.4583 = 11/24

*** Method 2**. Linear programming

Min $z = v_1 + v_2 + v_3 + v_4$

(1) $v_1 - (v_2/2 + v_3/6 + v_4/12) \geq 0$	(Continue when s_1)
(2) $v_2 - (v_3/3 + v_4/6) \geq 0$	(Continue when s_2)
(3) $v_3 - (v_4/4) \geq 0$	(Continue when s_3)
(4) $v_4 \geq 0$	(Continue when s_4)
(5) $v_1 \geq 0.25$	(Stop when s_1)
(6) $v_2 \geq 0.50$	(Stop when s_2)
(7) $v_3 \geq 0.75$	(Stop when s_3)
(8) $v_4 \geq 1.00$	(Stop when s_4)

and all variables v_j are *unrestricted in sign*.

- LP solution from Microsoft Excel - Solver

$$\mathbf{v}^* = [0.4583, 0.5000, 0.7500, 1.0000]$$

which agree with those found via the value iteration method.

- Optimal policy

a_1: Continue
The shadow price of (1) has a *non-zero* value.
The shadow prices of (2), (3), (4) are zero.

a_2: Stop
The shadow price of (5) is zero.
The shadow prices of (6), (7), (8) have *non-zero* values.

Thus, the optimal policy is to *continue* the search process (a_2) at the first stage (s_1), and then *stop* with a *candidate* thereafter. The success rate is 45.83% if you follow the optimal selection strategy.