Chapter 6:
Discrete probabilistic modeling

Linear Regression
Discrete-time Markov chains
Modeling reliability

Linear Regression

- A statistical methodology for minimizing the sum of the squared deviations

The basic linear regression model is defined by

$$\hat{y}_i = ax_i + b \quad \text{for } i = 1, 2, \ldots, m \text{ data points}$$

Review:

$$\text{SSE} = \sum_{i=1}^{m} [y_i - (ax_i + b)]^2$$

$$a = \frac{m \sum x_i y_i - \sum x_i \sum y_i}{m \sum x_i^2 - (\sum x_i)^2}, \text{ the slope}$$

$$b = \frac{\sum x_i^2 \sum y_i - \sum x_i y_i \sum x_i}{m \sum x_i^2 - (\sum x_i)^2}, \text{ the intercept}$$

Linear Regression

Sum Square Error:

$$\mathrm{SSE} = \sum_{i=1}^{m} [y_i - (ax_i + b)]^2$$

total corrected sum of squares

$$\mathrm{SST} = \sum_{i=1}^{m} (y_i - \bar{y})^2$$

where $\bar{y}$ is the average of the y values for the data points (x_i, y_i)

regression sum of squares

$$\mathrm{SSR} = \mathrm{SST} - \mathrm{SSE}$$

coefficient of determination

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}$$

The closer the value of R^2 is to 1, the better the fit of the regression line model to the actual data.

Linear Regression: example

- Modeling the relation of volume and the diameter of Ponderosa Pine (kind of tree).
- Comparing 4 models using computer and applying linear regression.

$$V = 0.00431d^3$$

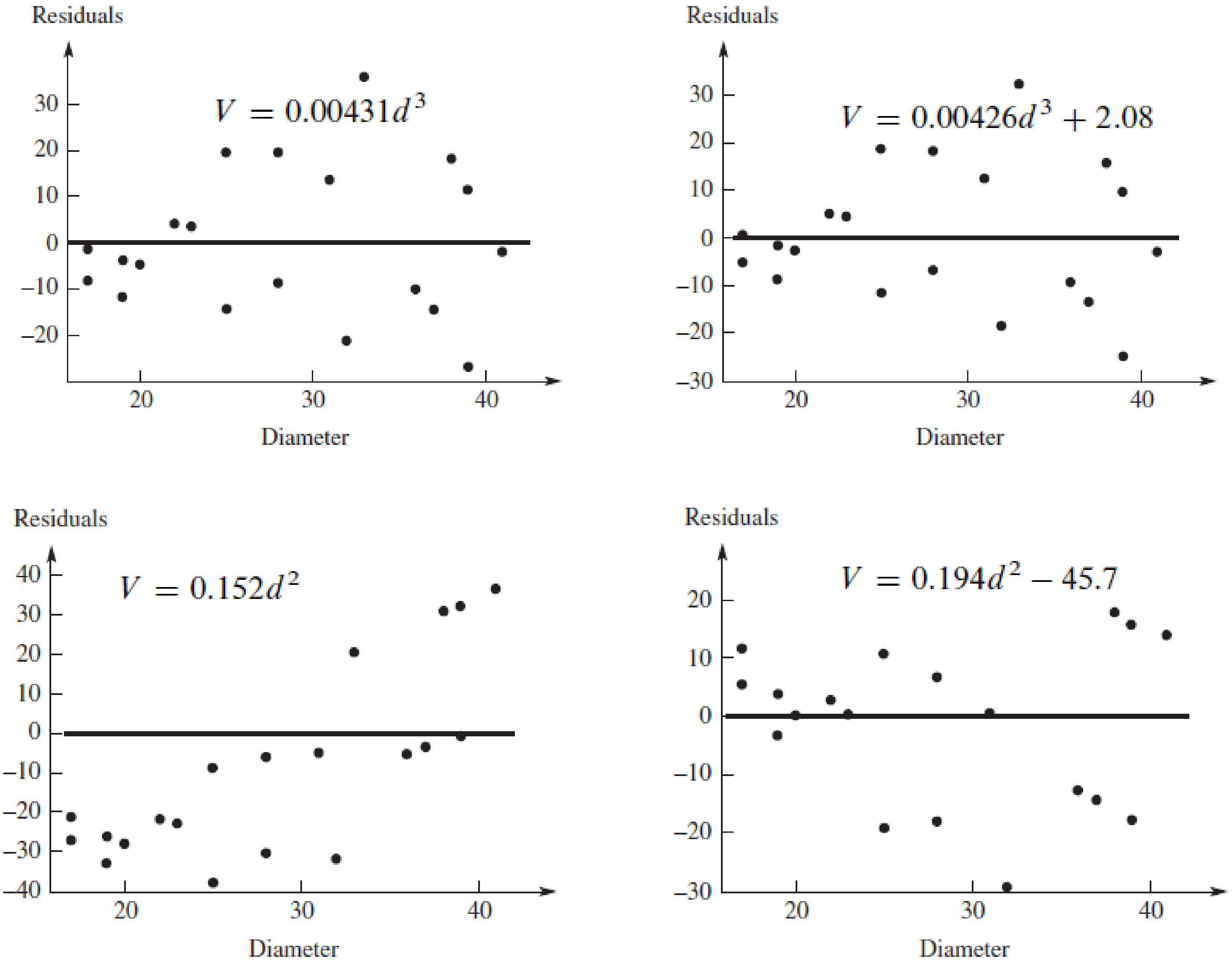
$$V = 0.00426d^3 + 2.08$$

$$V = 0.152d^2$$

$$V = 0.194d^2 - 45.7$$

Model	SSE	SSR	SST	R^2
$V = 0.00431d^3$	3,742	458,536	462,278	0.9919
$V = 0.00426d^3 + 2.08$	3,712	155,986	159,698	0.977
$V = 0.152d^2$	12,895	449,383	462,278	0.9721
$V = 0.194d^2 - 45.7$	3,910	155,788	159,698	0.976

Linear Regression: example



Worst model: A pattern in residuals

Probabilistic modeling

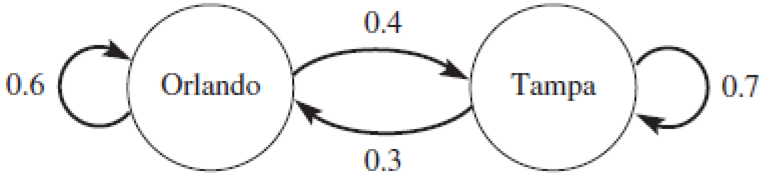
- So far in the course: deterministic models
  - precise sequences of actions, precise effects

- What if at every stage of the modeling there are several options of how to continue?
  - Non-deterministic or probabilistic models
  - Example: Orlando and Tamba rental car

transition matrix

Next state

		Orlando	Tampa
Present state	Orlando	0.6	0.4
	Tampa	0.3	0.7



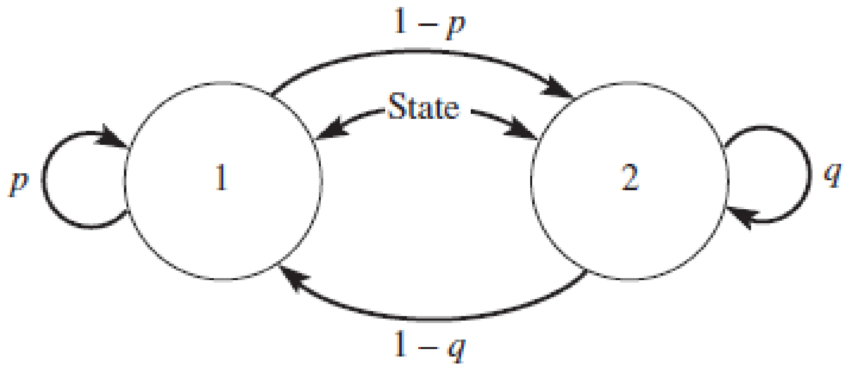
Markov Chain

- Markov chain is a process consisting of a sequence of events with the following properties:

1. The process is always in one of these states.
2. An event can transit the state to any other state or remain in the same state.
3. The probability of going from one state to another in a single stage is represented by a **transition matrix** for which the entries in each row are between 0 and 1; each row sums to 1.

Figure 6.1

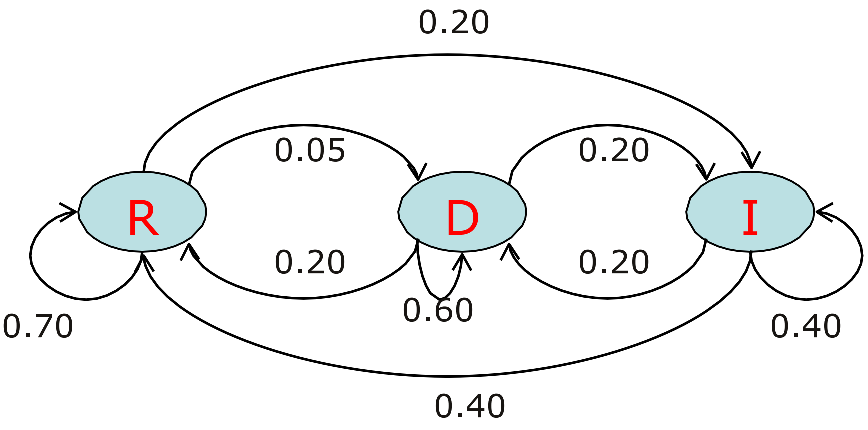
A Markov chain with two states; the sum of the probabilities for the transition from a present state is 1 for each state (e.g., $p + (1 - p) = 1$ for state 1).



Example: voting tendencies

- Assume 3 parties: Republicans, Democrats, Independents.
- Problem: identify the long-term behavior of voters in a presidential election
- Assumptions
    - ○ data collected over the last 10 years shows the following average trends in voting:

		Next election		
Last election		Republican	Democrat	Independent
	Republican	0.75	0.05	0.20
	Democrat	0.20	0.60	0.20
	Independent	0.40	0.20	0.40

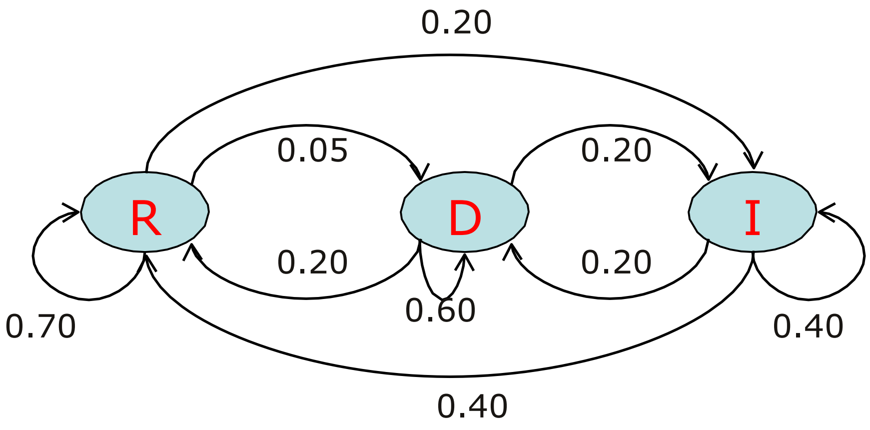


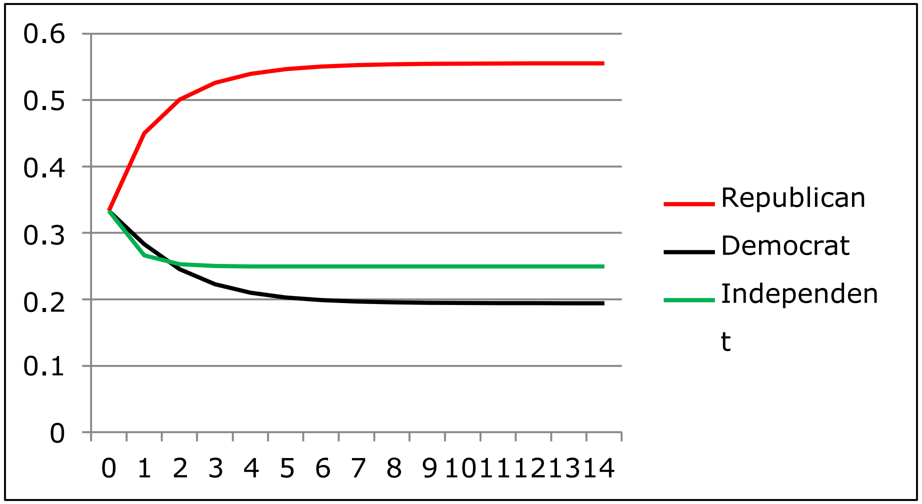
Example (continued)

- Model formulation
    - R_n=percentage of voters to vote Republican in election n
    - D_n=percentage of voters to vote Democratic in election n
    - I_n=percentage of voters to vote Independent in election n

$$R_{n+1} = 0.75R_n + 0.20D_n + 0.40I_n$$
$$D_{n+1} = 0.05R_n + 0.60D_n + 0.20I_n$$
$$I_{n+1} = 0.20R_n + 0.20D_n + 0.40I_n$$

- Numerical solution: start from an initial distribution of voters and calculate the model predictions for future elections and for long-term (asymptotic) behavior





Conclusions: discrete-time Markov chains

- Discrete state-space
  - The system is in a certain state at each step, the state changes randomly in the next step
  - Countable state-space
- Transitions
  - In each state there is a possible transition to any other state (including to itself)
  - The probability of each transition depends only on the current state, not on the sequence of events that led to the current step
  - Only one transition takes place in the current state, moving the system to its new state
    - in the new state, there will be a new probability distribution for the state transitions
- Discrete time
  - Steps can be defined in terms of time points, but also in terms of distance, number of events, or some other discrete measurement
  - Time advances in discrete “ticks” – the state is only updated at discrete time points

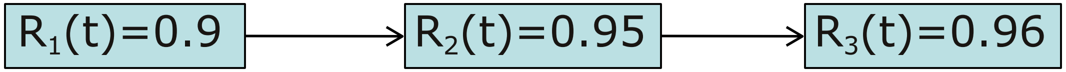
Conclusions: discrete-time Markov chains

- Graph representation
  - A Markov chain can be represented as a complete graph
  - The set of spaces is the set of nodes
  - Edges are marked with the probability of the corresponding transition
- State machine representation
  - A Markov chain can be seen as a computing device
  - Starts in an initial state, advances according to its transition table
  - It is a probabilistic machine
  - Unlike typical machines where the computation is expected to end with a final output, a Markov chain is rather expected to run an infinite computation
- Model checking
  - Various (qualitative and quantitative) questions may be asked
  - Reachability: is a certain state reachable from the initial state?
  - Is the probability of eventually reaching state s from the initial state 1?
  - What is the probability of a given property (say, reliability) after 100 steps?
  - …

Modeling Component and System Reliability

- A system is reliable if it performs well for a reasonably long time
  - reliability of a device: the probability that it will not fail over a specific time period

- **Series systems:** a sequence of devices that functions if and only if all of its devices are functioning

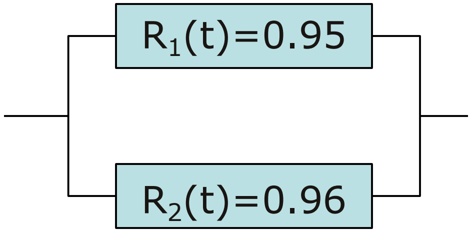


$R_1(t)=0.9$ → $R_2(t)=0.95$ → $R_3(t)=0.96$

  - The system's reliability: $R(t)=R_1(t)R_2(t)R_3(t)=0.8208$
  - less than each component's reliability!

Example (continued)

- **Parallel systems:** it functions well as long as at least one of its components functions well
  - in other words: it fails if and only if ALL of its devices fail



  - calculating the reliability R(t)
  - easier to reason in terms of failing 1-R(t)

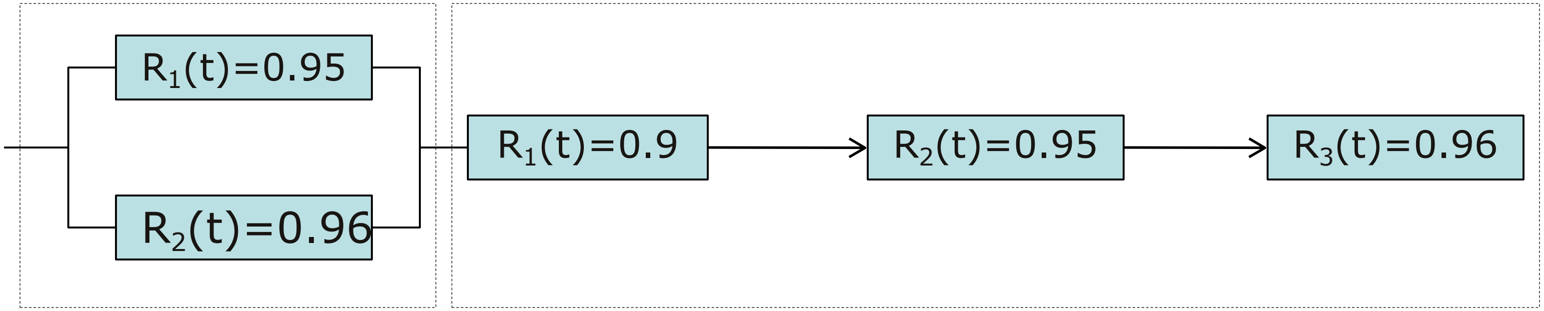
$$1-R(t)=(1-R_1(t))(1-R_2(t))$$

$$R(t)=R_1(t)+R_2(t)-R_1(t)R_2(t)=0.998$$

  - higher than any of the components' reliability

Example (continued)

- **A mix of series and parallel combinations:**



- Consider it as a series combination of the two systems on the previous slides
  - R(t)=R′(t)R″(t)=0.8192