1. Review of Probability

- Random Experiments
- The Axioms of Probability
- Conditional Probabilty
- Independence of Events
- Sequential Experiments
- Discrete-time Markov chain

Random Experiments

An experiment in which the outcome varies in an unpredictable fashion when the experiment is repeated under the same conditions

Examples:

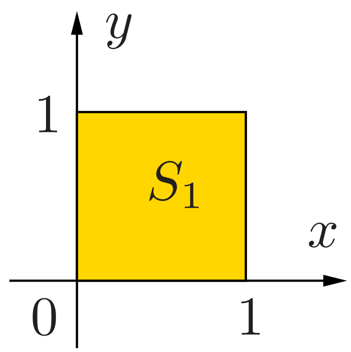
- Select a ball from an urn containing balls numbered 1 to n
- Toss a coin and note the outcome
- Roll a dice and note the outcome
- Measure the time between page requests in a Web server
- Pick a number at random between 0 and 1

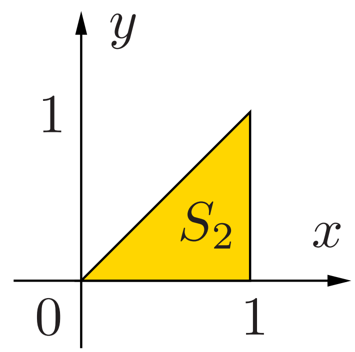
Sample space

Sample space is the set of all possible outcomes, denoted by S

- obtained by listing all the elements, e.g., $S = \{H, T\}$, or
- giving a property that specifies the elements, e.g., $S = \{x \mid 0 \leq x \leq 3\}$

Same experimental procedure may have different sample spaces





- Experiment 1: Pick two numbers at random between zero and one
- Experiment 2: Pick a number X at random between 0 and 1, then pick a number Y at random between 0 and X

Three possibilities for the number of outcomes in sample spaces

finite, countably infinite, uncountably infinite

Examples:

$$S_1 = \{1, 2, 3, \ldots, 10\}$$
$$S_2 = \{\text{HH,HT,TT,TH}\}$$
$$S_3 = \{x \in \mathbb{Z} \mid 0 \le x \le 10\}$$
$$S_4 = \{1, 2, 3, \ldots\}$$
$$S_5 = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid 0 \le y \le x \le 1\}$$
$$S_6 = \text{Set of functions } X(t) \text{ for which } X(t) = 0 \text{ for } t \ge t_0$$

Discrete sample space: if S is countable (S_1, S_2, S_3, S_4)

Continuous sample space: if S is not countable (S_5, S_6)

Events

Event is a subset of a sample space when the outcome satisfies certain conditions

Examples: A_k denotes an event corresponding to the experiment E_k

E_1 : Select a ball from an urn containing balls numbered 1 to 10

A_1 : An even-numbered ball (from 1 to 10) is selected

$$S_1 = \{1, 2, 3, \dots, 10\}, \quad A_1 = \{2, 4, 6, 8, 10\}$$

E_2 : Toss a coin twice and note the sequence of heads and tails

A_2 : The two tosses give the same outcome

$$S_2 = \{\text{HH,HT,TT,TH}\}, \quad A_2 = \{\text{HH,TT}\}$$

E_3 : Count # of voice packets containing only silence from 10 speakers

A_3 : No active packets are produced

$$S_3 = \{x \in \mathbb{Z} \mid 0 \leq x \leq 10\}, \quad A_3 = \{0\}$$

Two events of special interest:

- **Certain event**, S, which consists of all outcomes and hence always occurs
- **Impossible event** or **null event**, $\emptyset$, which contains no outcomes and never occurs

Review of Set Theory

- $A = B$ if and only if $A \subset B$ and $B \subset A$
- $A \cup B$ (union): set of outcomes that are in A *or* in B
- $A \cap B$ (intersection): set of outcomes that are in A *and* in B
- A and B are *disjoint* or *mutually exclusive* if $A \cap B = \emptyset$
- A^{c} (complement): set of all elements not in A
- $A \cup B = B \cup A$ and $A \cap B = B \cap A$
- $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- DeMorgan's Rules

$$(A \cup B)^{\mathsf{c}} = A^{\mathsf{c}} \cap B^{\mathsf{c}}, \quad (A \cap B)^{\mathsf{c}} = A^{\mathsf{c}} \cup B^{\mathsf{c}}$$

Axioms of Probability

Probabilities are numbers assigned to events indicating how likely it is that the events will occur

A *Probability law* is a rule that assigns a number $P(A)$ to each event A

$P(A)$ is called the *the probability of* A and satisfies the following axioms

Axiom 1 $P(A) \geq 0$

Axiom 2 $P(S) = 1$

Axiom 3 If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

Probability Facts

- $P(A^{\mathsf{c}}) = 1 - P(A)$
- $P(A) \leq 1$
- $P(\emptyset) = 0$
- If $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive then

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If $A \subset B$ then $P(A) \leq P(B)$

Conditional Probability

The probability of event A given that event B has occured

The conditional probability, $P(A|B)$, is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{for } P(B) > 0$$

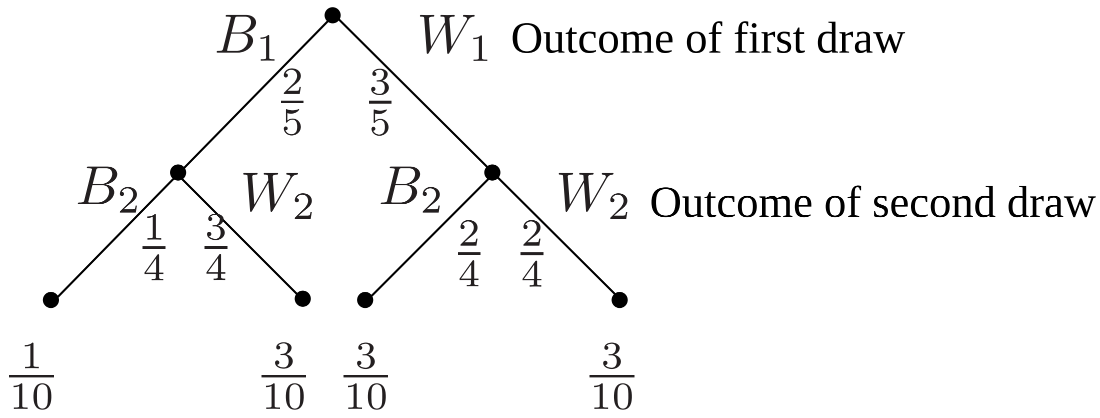
If B is known to have occured, then A can occurs only if $A \cap B$ occurs

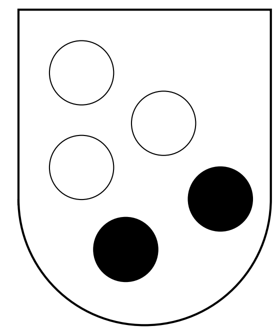
Simply renormalizes the probability of events that occur jointly with B

Useful in finding probabilities in sequential experiments

Example: Tree diagram of picking balls

Selecting two balls at random without replacement





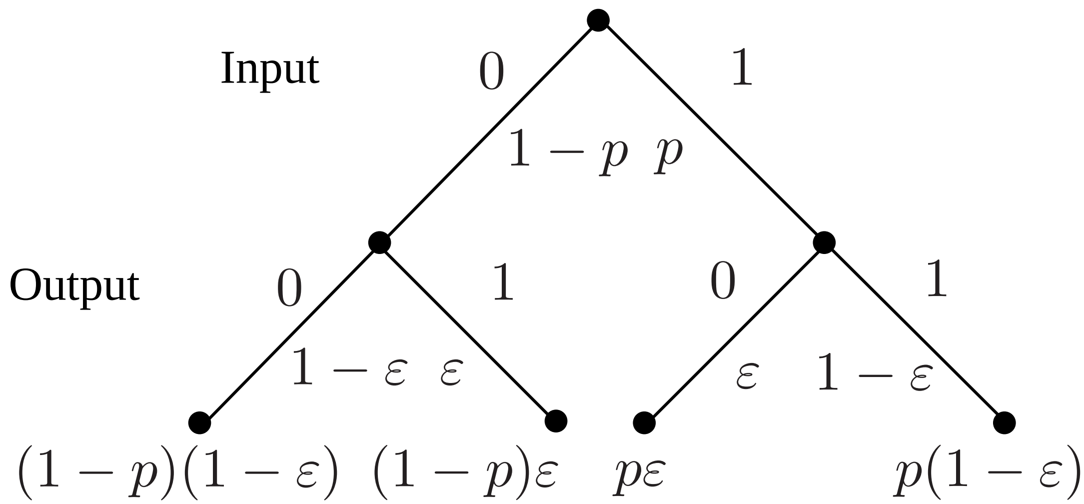
B_1, B_2 are the events of getting a black ball in the first and second draw

$$P(B_2|B_1) = \frac{1}{4}, \quad P(W_2|B_1) = \frac{3}{4}, \quad P(B_2|W_1) = \frac{2}{4}, \quad P(W_2|W_1) = \frac{2}{4}$$

The probability of a path is the *product* of the probabilities in the transition

$$P(B_1 \cap B_2) = P(B_2|B_1)P(B_1) = \frac{1}{4}\frac{2}{5} = \frac{1}{10}$$

Example: Tree diagram of Binary Communication



A_i: event the input was i, B_i: event the reciever was i

$$
\begin{aligned}
P(A_0 \cap B_0) &= (1-p)(1-\varepsilon) \\
P(A_0 \cap B_1) &= (1-p)\varepsilon \\
P(A_1 \cap B_0) &= p\varepsilon \\
P(A_1 \cap B_1) &= p(1-\varepsilon)
\end{aligned}
$$

Theorem on Total Probability

Let $B_1, B_2, \ldots, B_n$ be mutually exclusive events such that

$$S = B_1 \cup B_2 \cup \cdots \cup B_n$$

(their union equals the sample space)

Event A can be partitioned as

$$A = A \cap S = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)$$

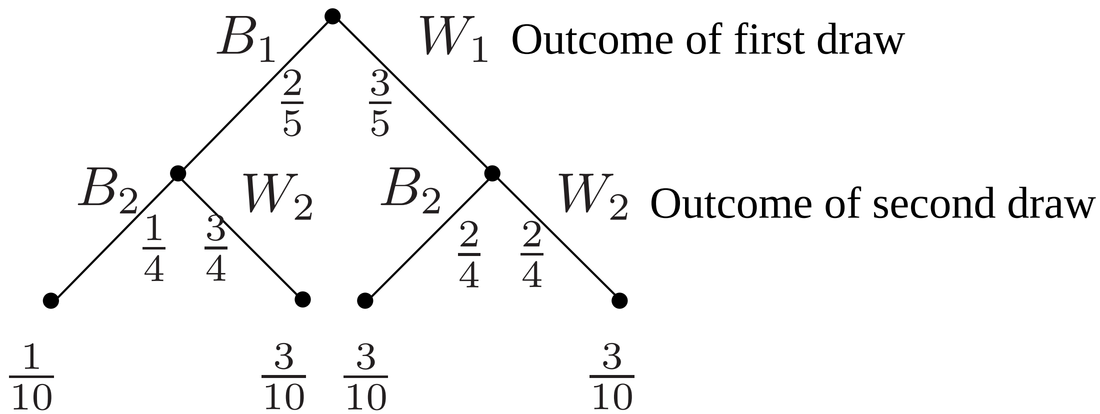
Since $A \cap B_k$ are disjoint, the probability of A is

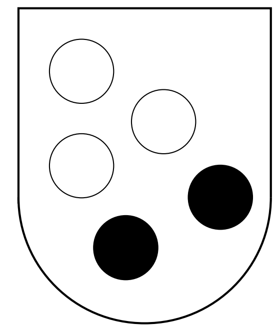
$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)$$

or equivalently,

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)$$

Example: revisit the tree diagram of picking two balls





Find the probability of the event that the second ball is white

$$
\begin{aligned}
P(W_2) &= P(W_2|B_1)P(B_1) + P(W_2|W_1)P(W_1) \\
&= \frac{3}{4} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{3}{5} = \frac{3}{5}
\end{aligned}
$$

Bayes' Rule

The conditional probablity of event A given B is related to the inverse conditional probability of event B given A by

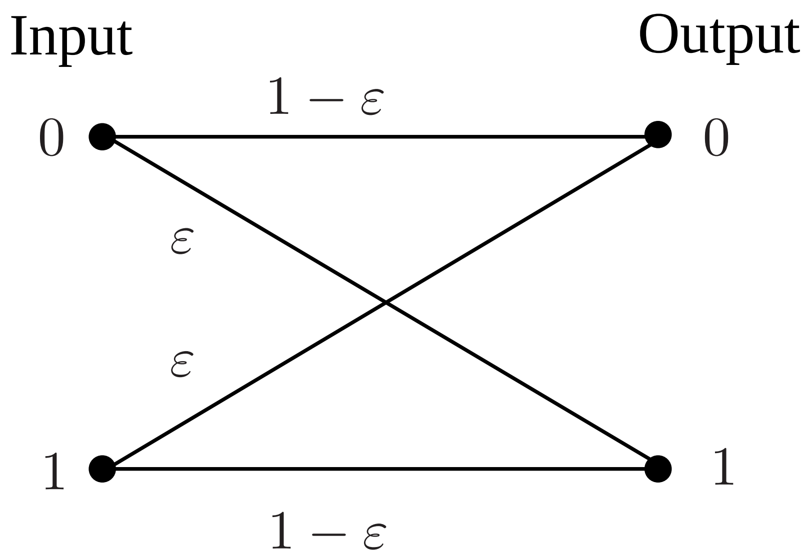
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- $P(A)$ is called a *priori* probability
- $P(A|B)$ is called a *posteriori* probability

Let $A_1, A_2, \ldots, A_n$ be a partition of S

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{n} P(B|A_k)P(A_k)}$$

Example: Binary Channel



A_i event the input was i

B_i event the receiver output was i

Input is equally likely to be 0 or 1

$$P(B_1) = P(B_1|A_0)P(A_0)+P(B_1|A_1)P(A_1) = \varepsilon(1/2)+(1-\varepsilon)(1/2) = 1/2$$

Applying Bayes' Rule, we obtain

$$P(A_0|B_1) = \frac{P(B_1|A_0)P(A_0)}{P(B_1)} = \frac{\varepsilon/2}{1/2} = \varepsilon$$

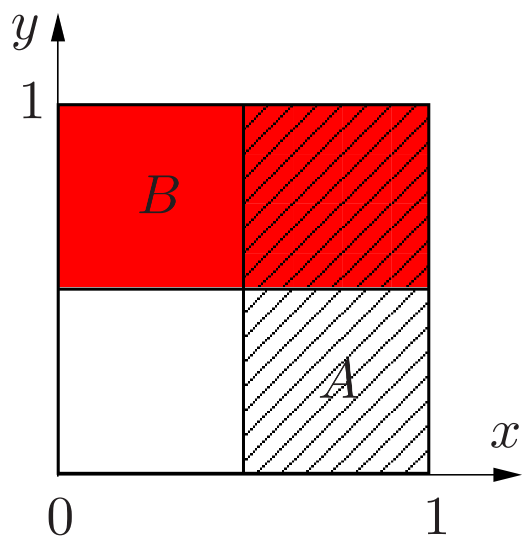
If $\varepsilon < 1/2$, input 1 is more likely than 0 when 1 is observed

Independence of Events

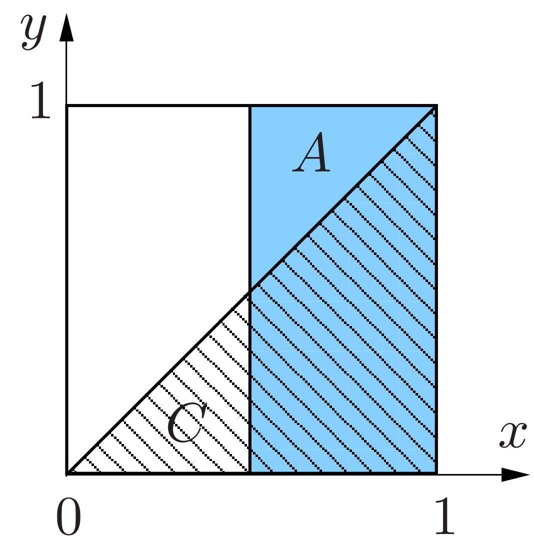
Events A and B are *independent* if

$$P(A \cap B) = P(A)P(B)$$

- Knowledge of event B does not alter the probability of event A
- This implies $P(A|B) = P(A)$

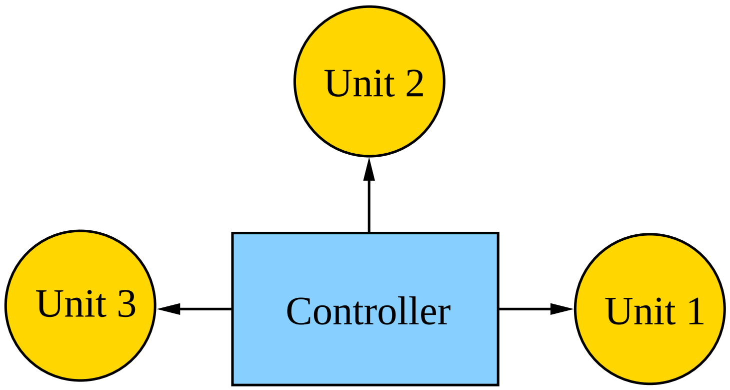


A and B are independent



A and C are not independent

Example: System Reliability



- System is 'up' if the controller and at least *two* units are functioning
- Controller fails with probability p
- Peripheral unit fails with probability a
- All components fail independently

- A: event the controller is functioning
- B_i: event unit i is functioning
- F: event two or more peripheral units are functioning

Find the probability that the system is up

The event F can be partition as

$$F = (B_1 \cap B_2 \cap B_3^c) \cup (B_1 \cap B_2^c \cap B_3) \cup (B_1^c \cap B_2 \cap B_3) \cup (B_1 \cap B_2 \cap B_3)$$

Thus,

$$\begin{aligned} P(F) &= P(B_1)P(B_2)P(B_3^c) + P(B_1)P(B_2^c)P(B_3) \\ &\quad + P(B_1^c)P(B_2)P(B_3) + P(B_1)P(B_2)P(B_3) \\ &= 3(1-a)^2 a + (1-a)^3 \end{aligned}$$

$$\begin{aligned} P(\text{system is up}) &= P(A \cap F) = P(A)P(F) \\ &= (1-p)P(F) = (1-p)\{3(1-a)^2 a + (1-a)^3\} \end{aligned}$$

Sequential Independent Experiments

- Consider a random experiment consisting of n *independent* experiments
- Let $A_1, A_2, \ldots, A_n$ be events of the experiments
- We can compute the probability of events of the sequential experiment

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\ldots P(A_n)$$

- Example: Bernoulli trial
  - Perform an experiment and note if the event A occurs
  - The outcome is "success" or "failure"
  - The probability of success is p and failure is $1 - p$

Binomial Probability

- Perform n Bernoulli trials and observe the number of successes
- Let X be the number of successes in n trials
- The probability of X is given by the *Binomial probability law*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

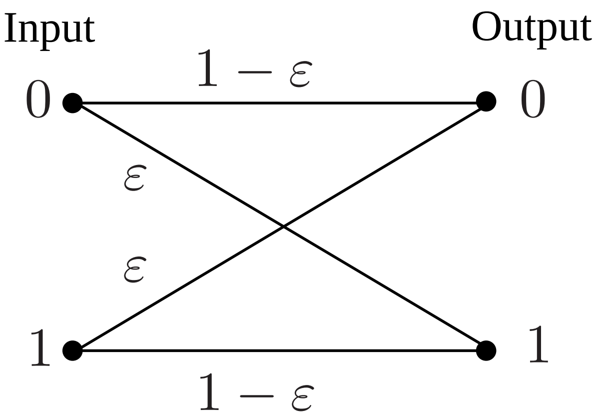
for $k = 0, 1, \ldots, n$

- The binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the number of ways of picking k out of n for the successes

Example: Error Correction Coding



- Transmit each bit three times
- Decoder takes a majority vote of the received bits

Compute the probability that the receiver makes an incorrect decision

- View each transmission as a Bernoulli trial
- Let X be the number of wrong bits from the receiver

$$P(X \geq 2) = \binom{3}{2} \varepsilon^2 (1-\varepsilon) + \binom{3}{3} \varepsilon^3$$

Mutinomial Probability

- Generalize the binomial probability law to the occurrence of more than one event
- Let $B_1, B_2, \dots, B_m$ be possible events with

$$P(B_k) = p_k, \quad \text{and} \quad p_1 + p_2 + \cdots + p_m = 1$$

- Suppose n independent repetitions of the experiment are performed
- Let X_j be the number of times each B_j occurs
- The probability of the vector $(X_1, X_2, \dots, X_m)$ is given by

$$P(X_1 = k_1, X_2 = k_2, \dots, X_m = k_m) = \frac{n!}{k_1!k_2!\dots k_m!} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$$

where $k_1 + k_2 + \cdots + k_m = n$

Geometric Probability

- Repeat independent Bernoulli trials until the the first success occurs
- Let X be the number of trials until the occurrence of the first success
- The probability of this event is called the *geometric probability law*

$$P(X = k) = (1-p)^{k-1}p, \quad \text{for } k = 1, 2, \ldots$$

- The geometric probabilities sum to 1:

$$\sum_{k=1}^{\infty} P(X = k) = p \sum_{k=1}^{\infty} q^{k-1} = \frac{p}{1-q} = 1$$

  where $q = 1 - p$

- The probability that more than n trials are required before a success

$$P(X > n) = (1-p)^n$$

Example: Error Control by Retransmission

- A sends a message to B over a radio link
- B can detect if the messages have errors
- The probability of transmission error is q
- Find the probability that a message needs to be transmitted more than two times

Each transmission is a Bernoulli trial with probability of success $p = 1 - q$

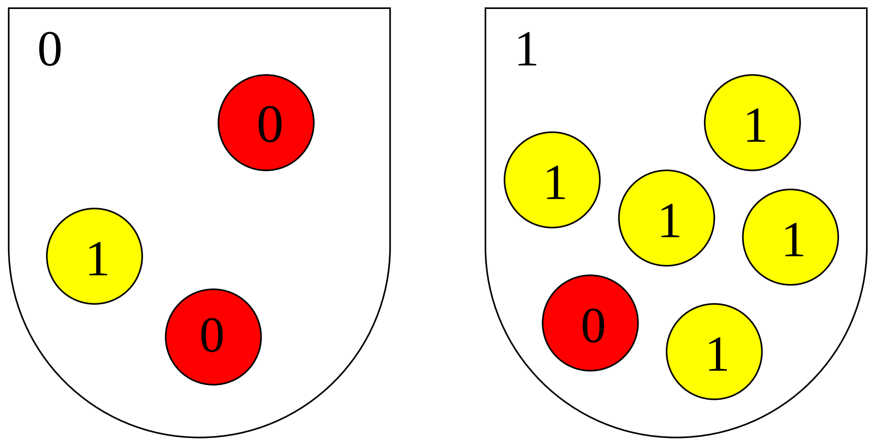
The probability that more than 2 transmissions are required is

$$P(X > 2) = q^2$$

Sequential Dependent Experiments

Sequence of subexperiments in which the outcome of a given subexperiment determine which subexperiment is performed next

Example: Select the urn for the first draw by flipping a fair coin

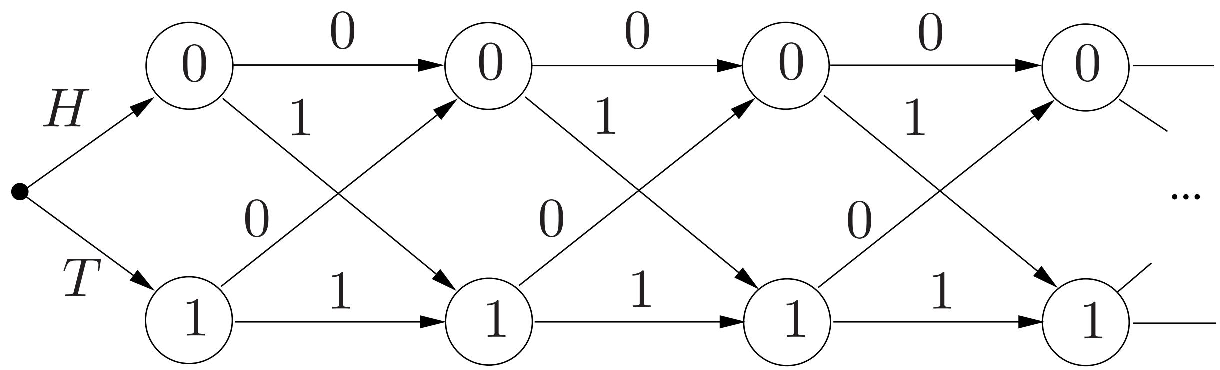


Draw a ball, note the number on the ball and replace it back in its urn

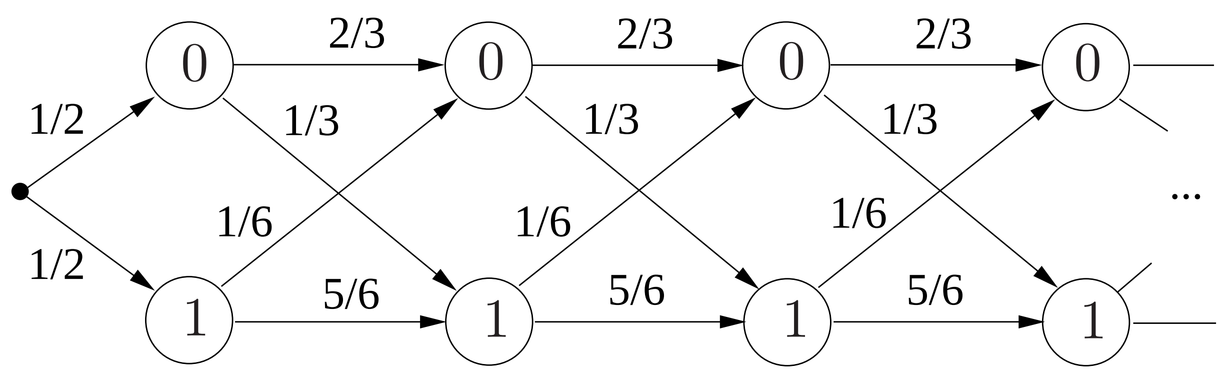
The urn used in the next experiment depends on # of the ball selected

Trellis Diagram

Sequence of outcomes



Probability of a sequence of outcomes



is the product of probabilities along the path

Markov Chains

Let $A_1, A_2, \ldots, A_n$ be a sequence of events from n sequential experiments

The probability of a sequence of events is given by

$$P(A_1 A_2 \cdots A_n) = P(A_n | A_1 A_2 \cdots A_{n-1}) P(A_1 A_2 \cdots A_{n-1})$$

If the outcome of A_{n-1} only determines the n^{th} experiment and A_n then

$$P(A_n | A_1 A_2 \cdots A_{n-1}) = P(A_n | A_{n-1})$$

and the sequential experiments are called *Markov Chains*

Thus,

$$P(A_1 A_2 \cdots A_n) = P(A_n | A_{n-1}) P(A_{n-1} | A_{n-2}) \cdots P(A_2 | A_1) P(A_1)$$

Find $P(0011)$ in the urn example

The probability of the sequence 0011 is given by

$$P(0011) = P(1|1)P(1|0)P(0|0)P(0)$$

where the transition probabilities are

$$P(1|1) = \frac{5}{6}, \quad P(1|0) = \frac{1}{3}, \quad P(0|0) = \frac{2}{3}$$

and the initial probability is given by

$$P(0) = \frac{1}{2}$$

Hence,

$$P(0011) = \frac{5}{6} \cdot \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{5}{54}$$

Discrete-time Markov chain

a Markov chain is a random sequence that has n possible states:

$$x(t) \in \{1, 2, \dots, n\}$$

with the property that

$$\mathbf{prob}(\ x(t+1) = i \ \mid \ x(t) = j\ ) = p_{ij}$$

where $P = [p_{ij}] \in \mathbf{R}^{n \times n}$

- p_{ij} is the **transition probability** from state j to state i
- P is called the **transition matrix** of the Markov chain
- the state $x(t)$ still cannot be determined with *certainty*

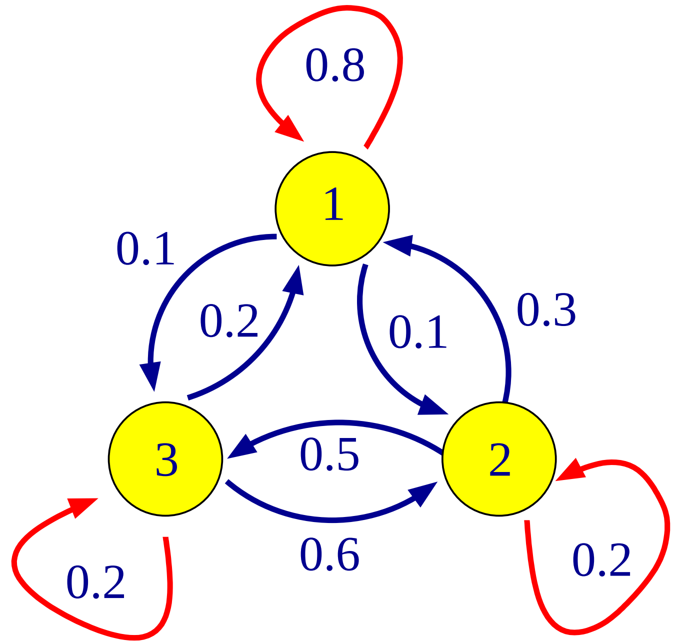
example:

a customer may rent a car from any of three locations and return to any of the three locations

Rented from location

1	**2**	**3**	
0.8	0.3	0.2	**1**
0.1	0.2	0.6	**2**
0.1	0.5	0.2	**3**

Returned to location



Properties of transition matrix

let P be the transition matrix of a Markov chain

- all entries of P are real *nonnegative* numbers
- the entries in any column are summed to 1 or $\mathbf{1}^T P = \mathbf{1}^T$:

$$p_{1j} + p_{2j} + \cdots + p_{nj} = 1$$

  (a property of a **stochastic matrix**)

- 1 is an eigenvalue of P
- if q is an eigenvector of P corresponding to eigenvalue 1, then

$$P^k q = q, \quad \text{for any } k = 0, 1, 2, \ldots$$

Probability vector

we can represent probability distribution of $x(t)$ as n-vector

$$p(t) = \begin{bmatrix} \mathbf{prob}(\ x(t) = 1\ ) \\ \vdots \\ \mathbf{prob}(\ x(t) = n\ ) \end{bmatrix}$$

- $p(t)$ is called a **state probability vector** at time t
- $\sum_{i=1}^{n} p_i(t) = 1$ or $\mathbf{1}^T p(t) = 1$
- the state probability propagates like a linear system:

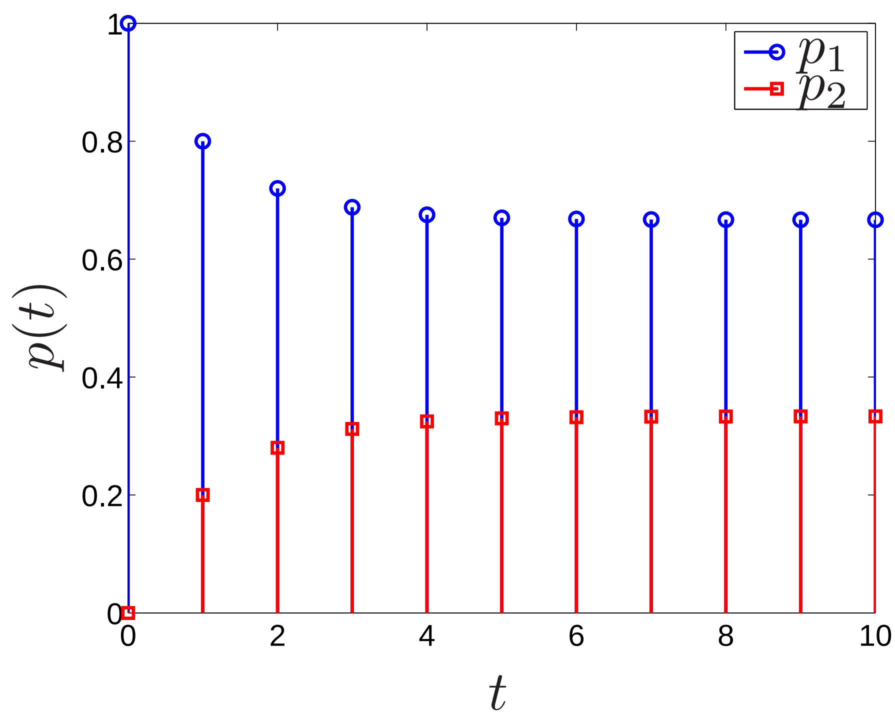
$$p(t+1) = Pp(t)$$

- the state PMF at time t is obtained by multiplying the initial PMF by P^t

$$p(t) = P^t p(0), \quad \text{for } t = 0, 1, \ldots$$

example: a Markov model for packet speech

- two states of packet speech: contain 'silent activity' or 'speech activity'
- the transition matrix is $P = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix}$
- the initial state probability is $p(0) = (1, 0)$
- the packet in the first state is 'silent' with certainty



- eigenvalues of P are 1 and 0.4
- calculate P^t by using 'diagonalization' or 'Cayley-Hamilton theorem'

$$P^t = \begin{bmatrix} (5/3)(0.4 + 0.2 \cdot 0.4^t) & (2/3)(1 - 0.4^t) \\ (1/3)(1 - 0.4^t) & (5/3)(0.2 + 0.4^{t+1}) \end{bmatrix}$$

- $P^t \rightarrow \begin{bmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{bmatrix}$ as $t \rightarrow \infty$ (all columns are the same in limit!)
- $\lim_{t\rightarrow\infty} p(t) = \begin{bmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} p_1(0) \\ 1 - p_1(0) \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$

$p(t)$ does not depend on the *initial state probability* as $t \rightarrow \infty$

what if $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$?

- we can see that

$$P^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad P^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ldots$$

- P^t does not converge but oscillates between two values

under what condition $p(t)$ converges to a constant vector as $t \to \infty$?

Definition: a transition matrix is **regular** if some integer power of it has all *positive* entries

Fact: if P is regular and let w be *any* probability vector, then

$$\lim_{t \to \infty} P^t w = q$$

where q is a **fixed** probability vector, independent of t

Steady state probabilities

we are interested in the **steady state probability vector**

$$q = \lim_{t \to \infty} p(t) \qquad \text{(if converges)}$$

- the steady-state vector q of a regular transition matrix P satisfies

$$\lim_{t \to \infty} p(t+1) = P \lim_{t \to \infty} p(t) \qquad \Longrightarrow \qquad Pq = q$$

(in other words, q is an eigenvector of P corresponding to eigenvalue 1)

- if we start with $p(0) = q$ then

$$p(t) = P^t p(0) = 1^t q = q, \quad \textbf{for all } t$$

q is also called the **stationary state PMF** of the Markov chain

example: weather model ('rainy' or 'sunny')

probabilities of weather conditions given the weather on the preceding day:

$$P = \begin{bmatrix} 0.4 & 0.2 \\ 0.6 & 0.8 \end{bmatrix}$$

(probability that it will rain tomorrow given today is sunny, is 0.2)

given today is sunny with probability 1, calculate the probability of a rainy day in long term

References

Chapter 2 in
A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition, Pearson Prentice Hall, 2009