Probability (Textbook Chapters 16 and 17)

R. Sekar

Monty Hall Problem

> *Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?*

Describes a situation faced by contestants on a 70's game show *Let's Make a Deal.*

Let's Make a Deal: Assumptions

- The car is equally likely to be hidden behind each of the three doors.
- The player is equally likely to pick each of the three doors.
- After the player picks a door, the host must open a different door with a goat behind it and offer the player a second choice.
- If the host has a choice of which door to open, then he is equally likely to select each of them.

The Sample Space

- *Random variables* (aka "random quantities")
  - door concealing the car.
  - door chosen by the player.
  - door opened by the host to reveal a goat.

  These variables take 3 possible values: A, B, and C, representing the three doors.

- *Outcome:* Values taken by random variables in any one experiment, e.g., (A, C, B) denotes:
  - the car is behind door A,
  - the player chooses door C,
  - the host opens door B

- *Sample space:* Set of all possible outcomes

$$S = \left\{ \begin{array}{c} (A,A,B),(A,A,C),(A,B,C),(A,C,B),(B,A,C),(B,B,A) \\ (B,B,C),(B,C,A),(C,A,B),(C,B,A),(C,C,A),(C,C,B) \end{array} \right\}$$

Events

A *set of outcomes* is called an event. Examples:

- "prize is behind door C"

$$\{(C, A, B), (C, B, A), (C, C, A), (C, C, B)\}$$

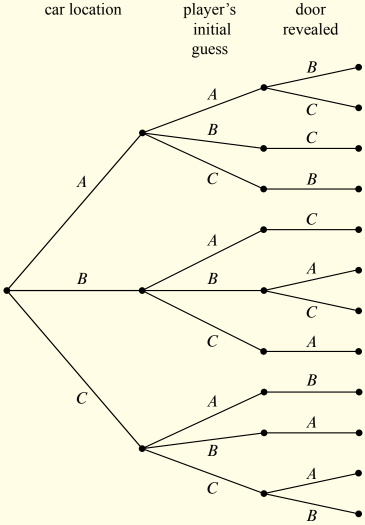
- "prize behind door first picked by the player"

$$\{(A, A, B), (A, A, C), (B, B, A), (B, B, C), (C, C, A), (C, C, B)\}$$

- "player wins by switching"

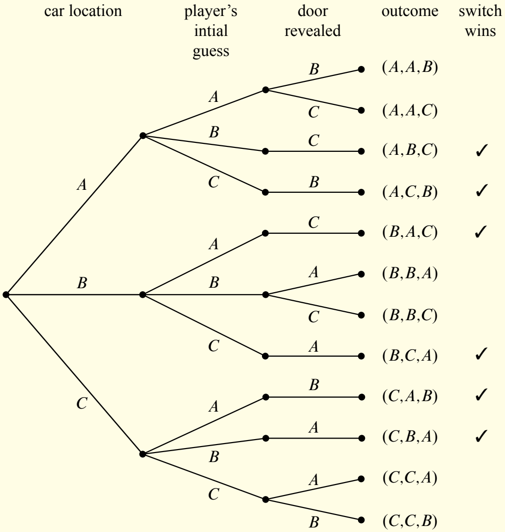
$$\{(A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)\}$$

Tree Diagram Displaying Sample Space

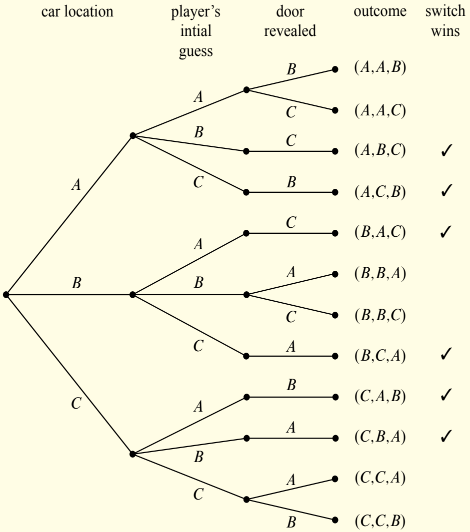


- Each *level* corresponds to a random variable
- Each *leaf* corresponds to an outcome
- Any *subset of leaves* correspond to an event

Tree Diagram With “Player Wins By Switching” Marked

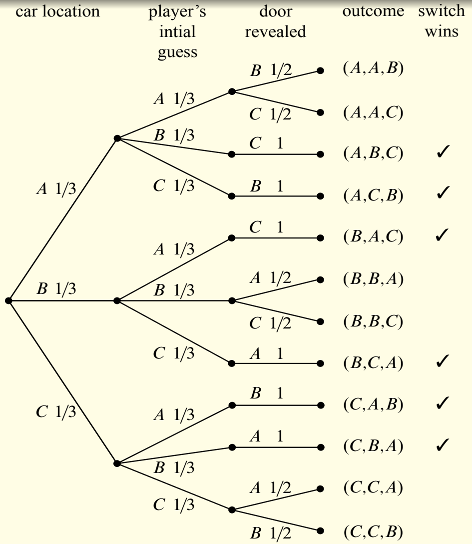


Computing Event Probabilities



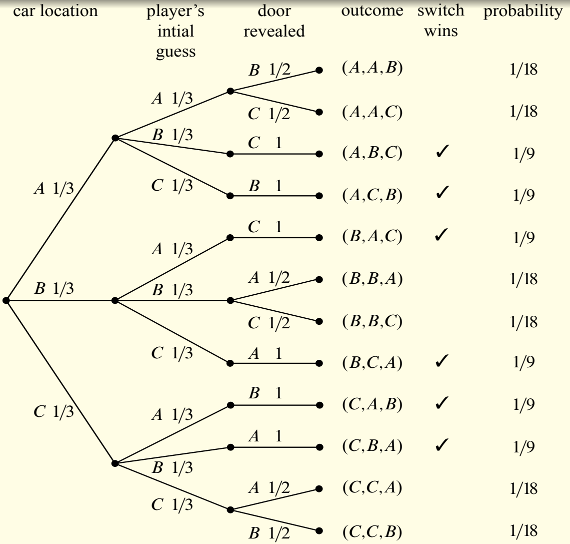
Computing Event Probabilities

- Assign edge probabilities



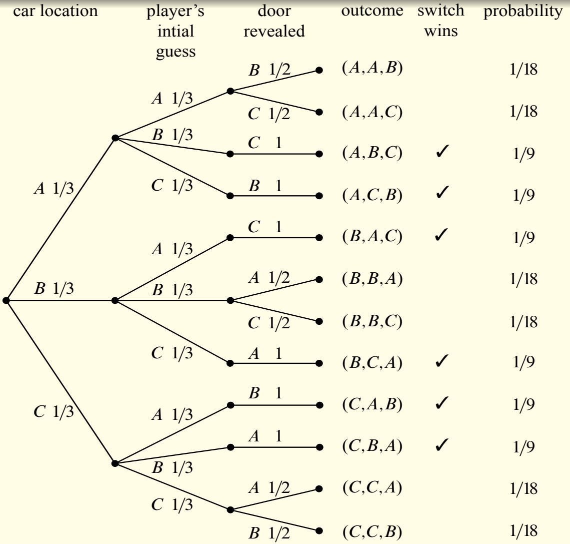
Computing Event Probabilities

- Assign edge probabilities
- Compute outcome probabilities



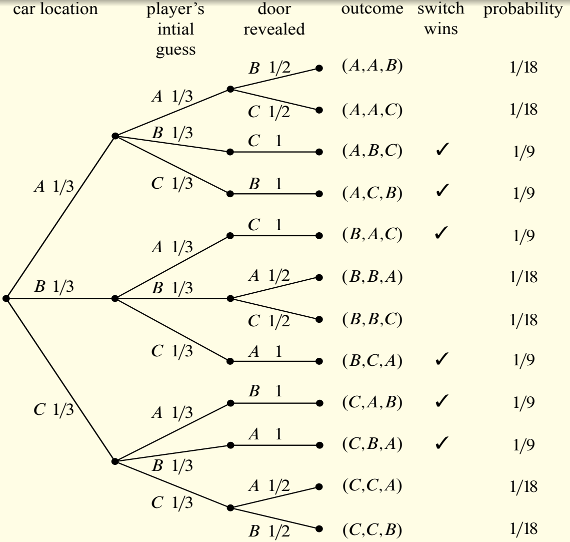
Computing Event Probabilities

- Assign edge probabilities
- Compute outcome probabilities
- Compute event probability:



Computing Event Probabilities

- Assign edge probabilities
- Compute outcome probabilities
- Compute event probability: $6/9 = 2/3$!



Birthday Problem

- What is the probability of finding two people with the same birthday in this class?
- The probability that two students have different birthdays: $\frac{364}{365}$
- In a class of n, there are $\binom{n}{2}$ pairs of students to consider.
  - If we assume that whether one pair shares a birthday is independent of another, we can simply multiply these probabilities

$$Pr(\text{no two persons with same birthday}) \approx \left(\frac{364}{365}\right)^{\binom{n}{2}} \approx \left(\frac{364}{365}\right)^{n^2/2}$$

- For $n = 44$, this formula yields a probability of 7%
  - $n = 23$ is enough to have better than even chance of finding two with the same birthday.

Birthday Problem: More Accurate Calculation

- What is the probability of finding two people with the same birthday in this class?
- Let us compute the complement: probability of *no* duplicated birthday, *assuming:*
  - birthdays are uniformly distributed in a year
  - birthdays of distinct students are independent.

$$\frac{365}{365} \cdot \frac{365-1}{365} \cdots \frac{365-(n-1)}{365} = \left(1 - \frac{0}{365}\right)\left(1 - \frac{1}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right)$$

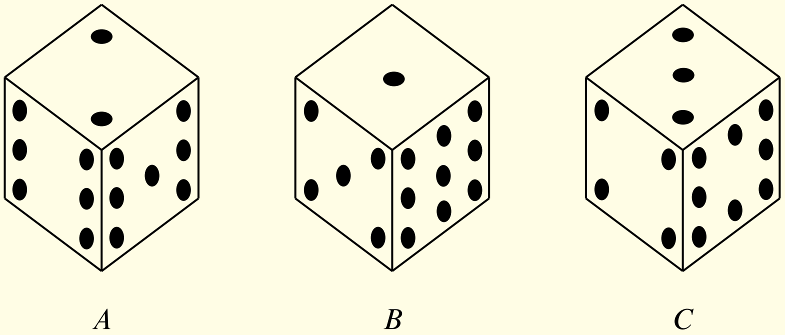
- Use the approximation $(1-x) < e^{-x}$ to derive an upper bound:

$$Pr(\text{no two persons with same birthday}) < e^0 \cdot e^{-\frac{1}{365}} \cdot e^{-\frac{n-1}{365}} = e^{\frac{-1}{365}\sum_{i=1}^{n-1} i} = e^{\frac{-n(n-1)}{2*365}}$$

- For $n = 44$, this evaluates to 7.5%

Strange Die Game

- A stranger challenges you to a game: whoever rolls higher will pay the other $10.



- To sweeten the deal, he says you can pick your die first.
- To keep you playing, he offers: sum of two rolls wins, *and*, he will go first!

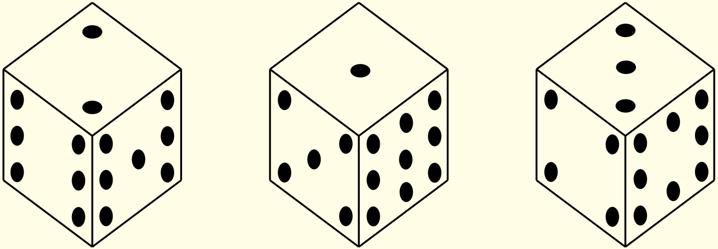
Strange Die Game: Summary

One roll game

- A beats B with a probability $5/9$
- B beats C with a probability $5/9$
- C beats A with a probability $5/9$

Sum of two rolls game

- A *loses to* B with a probability $42/81$
- B *loses to* C with a probability $42/81$
- C *loses to* A with a probability $42/81$



A B C

Set Theory and Probability

- A countable *sample space* $\mathcal{S}$ is a nonempty countable set.
- An *outcome* ω is an element of $\mathcal{S}$.
- A *probability function* $Pr : \mathcal{S} \longrightarrow \mathbb{R}$ is a total function such that
  - $Pr[\omega] \geq 0$ for all $\omega \in \mathcal{S}$, and
  - $\sum_{\omega \in \mathcal{S}} Pr[\omega] = 1$
- An *event* E is a subset of $\mathcal{S}$. Its probability is given by:

$$Pr[E] = \sum_{\omega \in E} Pr[\omega]$$

Probability Rules from Set Theory

Many probability rules follow from the rules on set cardinality

Sum Rule: If $E_0, E_1, \ldots, E_n, \ldots$ are pairwise disjoint events, then

$$Pr[\bigcup_{n \in \mathbb{N}} E_n] = \sum_{n \in \mathbb{N}} Pr[E_n]$$

Complement Rule: $Pr[\overline{A}] = 1 - Pr[A]$

Difference Rule:

$$Pr[B - A] = Pr[B] - Pr[A \cap B]$$

Inclusion–Exclusion:

$$Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$$

Union Bound: $Pr[A \cup B] \leq Pr[A] + Pr[B]$

Monotonicity: $A \subseteq B \rightarrow Pr[A] \leq Pr[B]$

Uniform Probability Spaces

A finite probability space $\mathcal{S}$ said to be uniform if $Pr[\omega]$ is the same for all ω. In such spaces:

$$Pr[E] = \frac{|E|}{|\mathcal{S}|}$$

We often this assumption — for instance, whenever probability was brought up while counting.

Infinite Probability Spaces

Two players take turns flipping fair coins. The first one to land heads wins. What is the probability of each player winning?

Conditional Probability

- Probability of an event under a condition
- The condition limits consideration to a subset of outcomes
  - Consider this subset (rather than whole of $\mathcal{S}$) as the space of all possible outcomes

$$Pr[X|Y] = \frac{Pr[X \cap Y]}{Pr[Y]}$$

Monty Hall Problem Revisited

$$Pr[X|Y] = \frac{Pr[X \cap Y]}{Pr[Y]}$$

Pr[win by switching | pick A and goat at B]

$Pr(\{(A,B,C),(A,C,B),(B,A,C),(B,C,A),(C,A,B),(C,B,A)\}|\{(A,A,B),(A,A,C),(C,A,B)\}]$
$= Pr[\{(C,A,B)\}]/Pr[\{(A,A,B),(A,A,C),(C,A,B)\}] = \frac{1/9}{1/18+1/18+1/9} = 1/2$

- Switching does not seem to help!

Monty Hall Problem Revisited

Wrong Question: $Pr[\text{win by switching} \mid \text{pick } A \text{ and goat at } B]$

$$Pr(\{(A,B,C),(A,C,B),(B,A,C),(B,C,A),(C,A,B),(C,B,A)\} | \{(A,A,B),(A,A,C),(C,A,B)\}]$$
$$= Pr[\{(C,A,B)\}]/Pr[\{(A,A,B),(A,A,C),(C,A,B)\}] = \frac{1/9}{1/18+1/18+1/9} = 1/2$$

- Switching does not seem to help!

Right Question: $Pr[\text{win by switching} \mid \text{pick } A \text{ and host opens } B]$

$$Pr(\{(A,B,C),(A,C,B),(B,A,C),(B,C,A),(C,A,B),(C,B,A)\} | \{(A,A,B),(C,A,B)\}]$$
$$= Pr[\{(C,A,B)\}]/Pr[\{(A,A,B),(C,A,B)\}] = \frac{1/9}{1/18+1/9} = 2/3$$

- Switching does help: The main clue is the host's decision to open B!

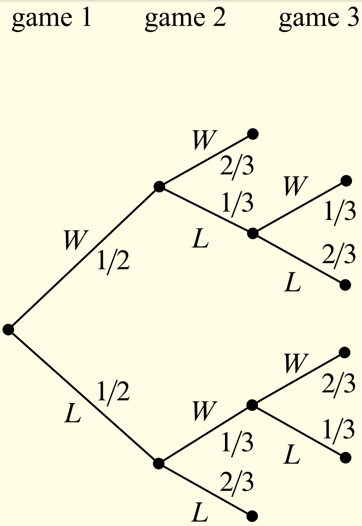
Four-Step Method for Conditional Probability

Best-of-Three Playoff

Both teams have a 0.5 probability of winning the first match. But for subsequent games, the winning team has a 2/3 probability of winning the next match. Similarly, the losing team has a 2/3 probability of losing the next match.

What is the probability that the team that wins the first match will win the playoffs?

Four-Step Method for Conditional Probability



outcome	event A: win the series	event B: win game 1	outcome probability
WW	✓	✓	1/3
WLW	✓	✓	1/18
WLL		✓	1/9
LWW	✓		1/9
LWL			1/18
LL			1/3

Four-Step Method for Conditional Probability

- Find the sample space

$$\mathcal{S} = \{WW, WLW, WLL, LWW, LWL, LL\}$$

- Define events of interest

$$W_A = \{WW, WLW, LWW\} \qquad W_B = \{WW, WLW, WLL\}$$

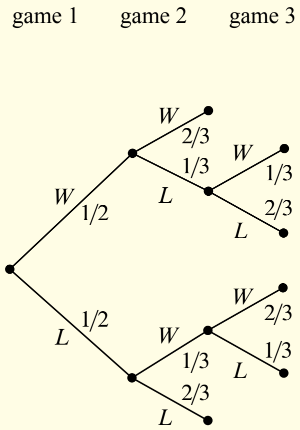
- Determine outcome probabilities
  - Outcomes correspond to the tree leaves, and are annotated with their probabilities
- Compute event probabilities

$$Pr[W_A] = \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{3} + \frac{1}{18} + \frac{1}{9} = \frac{1}{2}$$

$$Pr[W_A | W_B] = \frac{Pr[\{WW, WLW\}]}{Pr[W_B]} = \frac{1/3 + 1/18}{1/2} = \frac{7}{9}$$

What are Edge Probabilities in Tree Diagrams?

- They are just conditional probabilities!



outcome	event A: win the series	event B: win game 1	outcome probability
WW	✓	✓	1/3
WLW	✓	✓	1/18
WLL		✓	1/9
LWW	✓		1/9
LWL			1/18
LL			1/3

Extending Probability Rules for Conditional Probability

Product Rule 2: $Pr[E_1 \cap E_2] = Pr[E_1] \cdot Pr[E_2|E_1]$

Product Rule 3: $Pr[E_1 \cap E_2 \cap E_3] = Pr[E_1] \cdot Pr[E_2|E_1] \cdot Pr[E_3|E_1 \cap E_2]$

Bayes' Rule: $Pr[B|A] = \dfrac{Pr[A|B] \cdot Pr[B]}{Pr[A]}$

Total Probability Law: $Pr[A] = Pr[A|E] \cdot Pr[E] + Pr[A|\overline{E}] \cdot Pr[\overline{E}]$

Total Probability Law 2: If E_i are mutually disjoint and $Pr[\bigcup E_i] = 1$ then

$Pr[A] = \sum Pr[A|E_i] \cdot Pr[E_i]$

Inclusion-Exclusion: $Pr[A \cup B|C] = Pr[A|C] + Pr[B|C] - Pr[A \cap B|C]$

Independence

- An event A is independent of B iff the following (equivalent) conditions hold:
  - $Pr[A|B] = Pr[A]$
  - $Pr[A \cap B] = Pr[A] \cdot Pr[B]$
  - B is independent of A
- Often, independence is an assumption.
- Definition can be generalized to 3 (or n) events. Events E_1, E_2 and E_3a are mutually independent iff all of the following hold:
  - $Pr[E_1 \cap E_2] = Pr[E_1] \cdot Pr[E_2]$
  - $Pr[E_2 \cap E_3] = Pr[E_2] \cdot Pr[E_3]$
  - $Pr[E_1 \cap E_3] = Pr[E_1] \cdot Pr[E_3]$
  - $Pr[E_1 \cap E_2 \cap E_3] = Pr[E_1] \cdot Pr[E_2] \cdot Pr[E_3]$

Medical Testing

False Positive (FP): *Pr*[positive test | not sick]

In the context of statistical hypothesis testing:

- FP is called *type I error* or *significance* and denoted by the letter α
- $\gamma = 1 - \alpha$ is called *specificity* or *confidence* of the test.

False Negative: *Pr*[negative test | sick]

In statistical hypothesis testing,

- FN is called *type II error* and denoted β.
- $1 - \beta$ is called the *power* of the test.

Medical Testing

- Consider a diagnostic test with $FP = 0.05$ and $FN = 0.02$:
  - $Pr[\text{pos}|\neg\text{sick}] = 0.05$ $\qquad$ $Pr[\text{neg}|\text{sick}] = 0.02$
  - $Pr[\text{neg}|\neg\text{sick}] = 0.95$ $\qquad$ $Pr[\text{pos}|\text{sick}] = 0.98$
- If a test comes back positive, what is the likelihood that he/she has the disease?
- It depends ...
  - ... on what fraction of the tested population is actually sick.
  - Assume this is 1%.
    - i.e., $Pr[\text{sick}] = 0.01$

Medical Testing: Four-Step Method

- Find the sample space

$$\mathcal{S} = \{(sick, pos), (sick, neg), (\neg sick, pos), (\neg sick, neg)\}$$

- Determine outcome probabilities:
  - $Pr[(sick, pos)] = Pr[sick] \cdot Pr[pos|sick] = 0.01 \cdot 0.98 = .0098$
  - $Pr[(\neg sick, pos)] = Pr[\neg sick] \cdot Pr[pos|\neg sick] = 0.99 \cdot 0.05 = .0495$
- Define events of interest

$$Sick = \{(sick, pos), (sick, neg)\} \qquad Pos = \{(sick, pos), (\neg sick, pos)\}$$

- Compute conditional probability

$$Pr[Pos] = Pr[(sick, pos)] + Pr[(\neg sick, pos)] = .0098 + .0495 = 0.0593$$
$$Pr[sick|pos] = \frac{Pr[(sick, pos)]}{Pr[pos]} = 0.0098/0.0593 = 16.5\%$$

Although the test is more than 95% accurate, a positive does not mean much:
You have only a small (16.5%) chance of being actually sick!

Medical Testing: Summary

- While false positives are rare, they are more common that the likelihood of a random person being sick
  - In fact, the condition being tested is 5x less prevalent than FPs.
  - So, 4 out 5 times, people flagged by the test are not sick.
- This calculation is based on the assumption that the person being tested is someone picked randomly from the population.
  - If we tested only those that display symptoms of the sickness, the rates will be different.
    - In particular, we need to use the prevalence of sickness among such symptomatic people.