Time Waste Versus Empowerment

"I took a course in speed reading, and read War and Peace in 20 minutes. It's about Russia."

- Comedian Woody Allen

"I learned very early the difference between knowing the name of something and knowing something."

- Richard Feynman, Nobel laureate in physics

"The main goal [of this course] is self-actualization through the empowerment of claiming your education."

- Professor Marc Mangel

"Give me six hours to chop down a tree and I will spend the first four sharpening the axe."

- Abraham Lincoln

The Importance of Statistics

- Statistics impacts various aspects of daily life: business, medicine, law, government, etc.
- Examples:
  - Wall Street's models contributed to the 2008 financial crash.
  - Legal trials depend on understanding statistical evidence.
  - Probability in gambling and insurance.
  - Evaluating medical treatments.
  - Identifying potential terrorists.
- Understanding statistics is crucial for responsible decision-making.

Beyond Formulas

- Statistics is more than formulas.
- It can be mathematically deep and empowering.
- Everyone with calculus knowledge can understand statistics.
- Empowerment through understanding.

Mastering Probability

- Statistics is based on probabilistic models.
- Mastery of probability principles is essential for effective data analysis.
- Like "sharpening the axe" for effective problem-solving.

Embracing Understanding

- Focus on understanding equations and concepts.
- Connect mathematical ideas to real-life implications.
- Empowerment through comprehension.

"You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician."

Sherlock Holmes
The Sign of Four

Statistics and Data Science

Statistics is the science of collecting, classifying, summarizing, organizing, analyzing, and interpreting numerical or qualitative information based on data.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains.

Data Mining

Data mining is a process of extracting and discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

"To consult the statistician after an experiment is finished is often merely to ask him to perform a post-mortem examination. He can perhaps say what the experiment died of." *(R. A. Fisher)*

Statistics builds upon the foundation of probability.

Probability theory deals with mathematical formalizations of everyday concepts of chance, randomness, and related ideas.

This chapter introduces basic ideas of probability theory that are essential for statistics.

Basics of Probability

Definition
The **sample space** S for an experiment is the set of all possible outcomes, and outcomes are called **events**.

Note: A *set* can have finitely or infinitely many elements, and ordering and multiplicity are ignored.

Sample Space Examples

Example
Sample Space:

(i) Toss a coin once. $S =?$

(ii) Toss a coin twice. (a) Record outcomes regardless of order: $S =?$

(b) Record outcomes in order: $S =?$

(iii) (a) Number of rolls of a die until "1" appears. $S =?$

- (b) Experiment: Reaction time to a stimulus. $S =?$

Definition of Event

Definition
An **event** is a collection of possible outcomes of an experiment, i.e., any subset of S (including S itself).

An event A *occurs* if the experiment's outcome is in set A.

Interpretations of Probability:

- "Frequency of occurrence" of an event (Frequentist approach).
- Convergence of outcome frequencies to probability.
- *Subjective interpretation*: Probability as belief in event chance (Bayesian approach).
- *Function approach*: Probabilities defined by a function satisfying rules.

Axioms of Probability

Definition
A *probability* function P on sample space S satisfies:

A1. $P(A) \geq 0$ for all $A \subseteq S$.

A2. $P(S) = 1$.

A3. Countable Additivity: $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ for pairwise disjoint A_i.

Computing Probability

Result
If P is a probability function and A is any set in S, then:

(a) $P(\emptyset) = 0$,

(b) $P(A) \leq 1$,

(c) $P(A^c) = 1 - P(A)$.

Computing Probability (Contd.)

Result
If P is a probability function and A and B are any two sets in S, then:

(a) $P(A \cap B^c) = P(A) - P(A \cap B)$,

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,

(c) *If* $A \subset B$ *then* $P(A) \leq P(B)$.

Theorem
Fundamental Theorem of Counting: *If a procedure has tasks that can be done in n_i ways, then the entire procedure can be done in $n_1 \times n_2 \times \cdots n_k$ ways.*

Counting techniques are valuable for finite sample spaces with equally likely outcomes.

Conclusion

- Probability theory forms the basis for understanding chance and randomness.
- Probability's interpretations, rules, and counting techniques are crucial for calculating probabilities and making informed decisions.

Example

NY State Lottery The NY State Lottery operates as follows:
1. A person may pick any six numbers from the numbers $1, 2, \ldots, 44$ for his/her ticket.
2. The winning number is decided by randomly selecting six numbers from the forty-four numbers. Find the number of ways to pick the first two numbers, when

1. Repetition not allowed: Find # of ways to pick the first two numbers from 1 to 44.
2. Repetition allowed: Find # of ways to pick the first two numbers from 1 to 44.

Counting Principles

- Picking with replacement vs. picking without replacement
- Ordered picks vs. unordered picks

	Without Replacement	With Replacement
Ordered		
Unordered		

Permutations

Definition
A permutation rearranges elements without losing, adding, or changing them, where order matters.

- Permutations of n objects: $n! = n \times (n-1) \times \ldots \times 2 \times 1$
- Permutations of r out of n objects: $P(n, r) = \frac{n!}{(n-r)!}$

Combinations

Definition
A combination selects r out of n objects without rearrangement.

- Combinations of n objects: $\binom{n}{r} = \frac{n!}{(n-r)! \cdot r!}$

Derivation: Let $C(n, r)$ be the number of ways to select r out of n items (in any order).
So,

Counting Examples

Example

Six 6th graders and four 5th graders took an exam. (i) How many different rankings of the ten students are possible?

(ii) If 5th graders are to be ranked among themselves and 6th graders among themselves, how many different rankings are possible?

(iii) From the 6th graders, in how many ways can you select and rank 3 students?

Counting Examples

Example
What is the probability that a four-digit PIN does not have repeated digits?

Example
Committee selection: 6 men, 9 women. Probability of committee with 3 men and 2 women?

Ex: NY State Lottery (continued)

In how many ways can you pick 6 numbers from 1 to 44 in the lottery game?

1. **ordered, without replacement:**

   $$P(44,6) = \frac{44!}{(44-6)!} = 44\times43\times42\times41\times40\times39 = 5,082,517,440$$

2. **ordered, with replacement:**

   $$44^6 = 7,256,313,856$$

3. **unordered, without replacement:**

   $$\binom{44}{6} = \frac{44!}{(44-6)!6!} = \frac{44\times43\times42\times41\times40\times39}{6\times5\times4\times3\times2\times1} = 7,059,052$$

4. **unordered, with replacement:** This is the most difficult case to count.

   $$\frac{(44+6-1)!}{(44-1)!6!} = 13,983,816$$

Ex: NY State Lottery (continued)

In summary, we have the following table:

Table 1: Number of possible arrangements of size r from n objects

	Without Replacement ($r \leq n$)	With Replacement
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

Birthday Problem

Suppose there are n ($n \leq 365$) students in a class. Assuming there are 365 days in a year, what is the probability that at least two students have the same birthday?

Solution: Define the event

$A = \{\text{At least two students have the same birthday}\}$,

then $A^c = \{\text{All students have different birthdays}\}$.

\# of elements in the sample space is $\#(S) = 365^n$,

\# of elements in the event is $\#(A^c) = P(365, n) = \dfrac{365!}{(365-n)!}$.

Then $P(A^c) = \dfrac{\left(\frac{365!}{(365-n)!}\right)}{365^n}$. So, the answer is

$$P(A) = 1 - P(A^c) = 1 - \frac{\left(\frac{365!}{(365-n)!}\right)}{365^n}.$$

Birthday Problem

n	10	20	30	40	50
$P(A)$	0.12	0.41	0.71	0.89	0.97

Table 2: The probability values $P(A)$ for various n (in the Birthday Problem).

Example

(Challenge Birthday Problem) In the same setting of the Birthday Problem, what is the probability that

(i) at least k students have the same birthday (for $2 < k \leq n$)?

(ii) exactly k students have the same birthday (for $2 \leq k \leq n$)?

Conditional Probability and Independence

In many instances, we update the sample space based on new information, leading to conditional probabilities.

Example
(Die Roll) Roll a fair die.

1. Find the probability of getting 1.
2. Find the probability of getting 1 when the number is odd.

Definition
(Conditional Probability) If A and B are events and $P(B) > 0$, then $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Examples

Example

(Two Coin Flips) A fair coin is flipped twice. Let E be both flips resulting in heads, and F be the first flip resulting in a head. Then $P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{1/4}{2/4} = \frac{1}{2}$.

Example

(Drawing Marbles) An urn contains 8 red and 4 white marbles. Draw 2 marbles without replacement.

$P(\text{Both red}) = \frac{8}{12} \times \frac{7}{11} = \frac{14}{33}$.

Law of Total Probability

Law of Total Probability: For A and B events,
$P(A) = P(A \cap B) + P(A \cap B^c) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$.

Generalization for mutually exclusive events:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i).$$

Bayes' Rule:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}.$$

Independence

Definition
(Independence) Two events, A and B, are statistically independent if $P(A \cap B) = P(A)P(B)$.

(i) It is tempting to say that events A, B and C are independent if $P(A \cap B \cap C) = P(A)P(B)P(C)$, but this is not correct!

(ii) Events A, B, and C are not mutually independent despite being pairwise independent.

Independence

(Example for (i): Rolling two fair dice) Roll two fair dice.
Then, the sample space is: $S = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(2,6),\ldots,(6,1),\ldots,(6,6)\}$.

Define the following events:
$A = \{\text{doubles occur}\} = \{\text{same \#s on the dice}\} = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$,

$$B = \{\text{sum is between 7 and 10 (inclusive)}\} = \{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3),(2,6),(6,2),(3,5),(5,3),(4,4),(3,6),(6,3),(4,5),(5,4),(4,6),(6,4),(5,5)\},$$

$$C = \{\text{sum is 2 or 7 or 8}\} = \{(1,1),(1,6),(6,1),(2,5),(5,2),(3,4),(4,3),(2,6),(6,2),(3,5),(5,3),(4,4)\},$$

Ex: (Rolling two fair dice, continued)

Then $A \cap B \cap C = \{(4,4)\}$, and

$$B \cap C = \{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3),(2,6),(6,2),\\ (3,5),(5,3),(4,4)\}.$$

Hence, $P(A) = 1/6$, $P(B) = 1/2$, $P(C) = 1/3$ and

$$P(A \cap B \cap C) = \frac{1}{36} = P(A) \cdot P(B) \cdot P(C) = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{3}$$

but

$$P(B \cap C) = 11/36 \neq P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{3}.$$

Thus, A, B, C are not mutually independent. □

It is also possible that events A, B and C are pairwise independent, but not mutually independent!

Conclusion

- Conditional probabilities allow us to update probabilities based on new information.
- Independence between events simplifies calculations and modeling.
- Care must be taken when assessing mutual independence.

Basic Probability Models

Introduction:
We continue to introduce fundamental probability concepts. While many of these ideas may seem intuitive, understanding them in detail is crucial, especially when dealing with complex scenarios where intuition alone might not suffice. The principles discussed here lay the groundwork for more advanced probability applications.

ALOHA Network Example

In this section, we discuss a computer network example. Probability analysis is instrumental in advancing faster network technologies.

- Today's Ethernet has its origins in ALOHA, an experimental network from the University of Hawaii.
- ALOHA featured multiple network nodes attempting to use a single radio channel for communication with a central computer. Because of geographical obstacles, the nodes couldn't hear each other.
- A single node's transmission succeeded and received an acknowledgment, but simultaneous transmissions led to collisions, rendering messages unintelligible.
- In case of collisions, sending nodes would timeout and reattempt later. To minimize collisions, nodes adopted random backoff, temporarily refraining from sending data.

ALOHA Network Example

- In this context, a variation called "slotted" ALOHA is introduced.
- This variation divides time into intervals referred to as "epochs," each lasting 1.0 unit. For instance, epoch 1 spans from time 0.0 to 1.0, epoch 2 from 1.0 to 2.0, and so on.
- Within each epoch, active nodes, meaning nodes with messages to send, make a decision to send or refrain with probabilities p and 1-p respectively.
- The designer sets the value of p, mimicking real Ethernet hardware's behavior that employs random number generators.

ALOHA Network Example

- Another parameter q in this model signifies the probability of an inactive node generating a message in an epoch, simulating the randomness of users hitting keys on a keyboard. This models the typical behavior of users intermittently typing on a computer.
- Let's assume a simple scenario with two nodes, $n = 2$, and for simplicity, message arrival coincides with the middle of an epoch. Decisions to send or back off occur approximately 90% into an epoch.
- In a given example, consider an epoch from time 15.0 to 16.0. At the epoch's beginning, node A has a message to send while node B does not.
- At time 15.5, node B makes a decision whether to generate a message to send or not, with probabilities q and 1-q respectively.

ALOHA Network Example

- Assuming B generates a message, at time 15.9, both node A and node B independently decide whether to send or refrain, each with probabilities p and 1-p.
- For this scenario, if A refrains but B sends, B's transmission succeeds, and at epoch 16, B becomes inactive, while A remains active. Conversely, if both A and B attempt to send at 15.9 and both fail, both nodes will still be active at epoch 16.
- Keep in mind that in this simplified model, when a node is active, it does not generate additional new messages during its active time.

ALOHA Network Model Summary

- We have n network nodes, sharing a common communications channel.
- Time is divided in epochs. X_k denotes the number of active nodes at the end of epoch k.
- If two or more nodes try to send in an epoch, they collide, and the message doesn't get through.
- We say a node is active if it has a message to send.
- If a node is active node near the end of an epoch, it tries to send with probability p; and If a node is inactive at the beginning of an epoch, then at the middle of the epoch it will generate a message to send with probability q.

Calculating $P(X_1 = 2)$

- Let's examine the network over two epochs: epoch 1 and epoch 2. We consider a network with only two nodes, referred to as node 1 and node 2. Initially, both nodes are active. Let X_1 and X_2 represent the counts of active nodes at the *end* of epochs 1 and 2, respectively, accounting for possible transmissions. For this example, we'll assume $p = 0.4$ and $q = 0.8$.
- Let's determine the probability that $X_1 = 2$ and subsequently delve into the underlying interpretation of this probability.
- How could $X_1 = 2$ happen? In our examples here, we have n $= 2$ and $X_0 = 2$, i.e. both nodes start out active.
- Two cases:
    - Both nodes send (p^2).
    - Both nodes refrain $(1 - p)^2$.
- $P(X_1 = 2) = p^2 + (1 - p)^2 = 0.52$.

Interpreting Probability: Practical Understanding

- Importance of understanding practical implications of probabilities.
- Example shift: ALOHA to dice rolling.
- Consider the "experiment" consisting of rolling two dice, say a blue one and a yellow one. Let X and Y denote the number of dots we get on the blue and yellow dice, respectively.
- Consider $P(X + Y = 6) = \frac{5}{36}$ as an illustration.

Dice Rolling: $P(X+Y=6)=\frac{5}{36}$

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Table 1: Sample Space for Dice Example

Dice Rolling: Theoretical Explanation

- Mathematical theory of probability involves a sample space.
- Weights on outcomes (1/36 each), e.g., (1,5), (2,4), (3,3), (4,2), (5,1).
- For example, "What we mean by $P(X + Y = 6) = \frac{5}{36}$ is that the outcomes (1,5), (2,4), (3,3), (4,2), (5,1) have total weight 5/36."
- Limitations: Complexity requires advanced math (measure theory).

Practical Probability Computation

- Most probability computations don't need explicit sample space.
- Sample space useful for explaining concepts.
- Intuition $>$ Theoretical grounding for understanding probabilities.

Loss of Intuition in Sample Space Approach

- Complex probability models lead to tricky sample space definitions.
- No effective way to convey intuition for conditional probability.
- Same limitation for expected value (central topic).

Intuitive Understanding of Probability

- Imagine repeating the experiment multiple times.
- Record outcomes in a notebook.
- Fraction of lines where event happens defines probability.
- Applying this idea simplifies probability problems.
- Fundamental basis of computer simulation.

Our Definitions: Intuition Over Rigor

- Intuitive definitions for practical understanding.
- Focusing on definitions, not properties.
- Experiment repeatability concept.

Events and Random Variables

- Events are possible boolean outcomes.
- Random variables are numerical or categorical outcomes.
- Experiment repetition and notebook context.
- Long-run fraction definition for probabilities.
- most probability computations do not rely on explicitly writing down a sample space.

Definitions and Properties

- Disjoint events and their probabilities.
- $$P(A \text{ or } B) = P(A) + P(B) \tag{1}$$

  for disjoint events.
- Independent events and their probabilities.
- $$P(A \text{ and } B) = P(A) \cdot P(B) \tag{2}$$

  for independent events.

- **Conditional Probability**
- Conditional probability $P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$.
- Independence simplifies $P(B|A) = P(B)$.

General Cases and Independence

- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ for general cases.
-
$$P(A \text{ and } B) = P(A)P(B|A) \tag{3}$$
  for non-independent events.
- Intuition in determining independence.
- Generalization of (2) to (3).

“Mailing Tubes” Strategy

Introducing the concept of using (2) and (3) as a strategy to calculate probabilities. These equations act like ”mailing tubes” that help compute probabilities by breaking down complex events into simpler ones.

- Using (2) and (3) as strategic tools.
- Applying the ”mailing tubes” metaphor to probability calculations.
- Emphasizing the importance of the strategy for simplifying complex probability problems.

Probabilities in Notebook Context

The intuitive notion—which is FAR more important—of what $P(X + Y = 6) = \frac{5}{36}$ means is the following. Imagine doing the experiment many, many times, recording the results in a large notebook:

- Roll the dice the first time, and write the outcome on the first line of the notebook.
- Roll the dice the second time, and write the outcome on the second line of the notebook.
- Roll the dice the third time, and write the outcome on the third line of the notebook.
- Roll the dice the fourth time, and write the outcome on the fourth line of the notebook.
- Imagine you keep doing this, thousands of times, filling thousands of lines in the notebook.

Probabilities in Notebook Context

notebook line	outcome	blue+yellow = 6?
1	blue 2, yellow 6	No
2	blue 3, yellow 1	No
3	blue 1, yellow 1	No
4	blue 4, yellow 2	Yes
5	blue 1, yellow 1	No
6	blue 3, yellow 4	No
7	blue 5, yellow 1	Yes
8	blue 3, yellow 6	No
9	blue 2, yellow 5	No

Table 2: Notebook for the Dice Problem

Probabilities in Notebook Context

- The first 9 lines of the notebook might look like Table 2. Here 2/9 of these lines say Yes.
- But after many, many repetitions, approximately 5/36 of the lines will say Yes.
- For example, after doing the experiment 720 times, approximately $\frac{5}{36} \times 720 = 100$ lines will say Yes.
- This is what probability really is: In what fraction of the lines does the event of interest happen? **It sounds simple, but if you always think about this "lines in the notebook" idea, probability problems are a lot easier to solve.**
- And it is the fundamental basis of computer simulation.

Probabilities in Notebook Context

- P(A) means the long-run fraction of lines in the notebook in which the A column says Yes.
- P(A or B) means the long-run fraction of lines in the notebook in which the A-or-B column says Yes.
- P(A and B) means the long-run fraction of lines in the notebook in which the A-and-B column says Yes.
- P(A | B) means the long-run fraction of lines in the notebook in which the A | B column says Yes—**among the lines which do NOT say NA.**

Confusing Probabilities: P(A and B) vs. P(A | B)

- P(A and B) and P(A | B) are distinct.
- Importance of the notebook view.
- Dice example: $P(X = 1 \text{ and } S = 6) = \frac{1}{36}$ vs. $P(X = 1|S = 6) = \frac{1}{5}$.

Example: ALOHA Network

$$P(X_1 = 2) = p^2 + (1-p)^2 = 0.52 \quad (4)$$

How did we get this?

- C_i denotes event node i tries to send ($i = 1, 2$).
- Using definitions and probability properties:

$$\begin{aligned}
P(X_1 = 2) &= P(\underbrace{C_1 \text{ and } C_2}\text{ or }\underbrace{\text{ not } C_1 \text{ and } \text{ not } C_2}) && (5)\\
\text{(from (1))} &= P(C_1 \text{ and } C_2) + P(\text{ not } C_1 \text{ and } \text{ not } C_2) && (6)\\
\text{(from (2))} &= P(C_1)P(C_2) + P(\text{ not } C_1)P(\text{ not } C_2) && (7)\\
&= p^2 + (1-p)^2 && (8)
\end{aligned}$$

Steps Explanation

- (5): List ways $\{X_1 = 2\}$ can occur.
- (6): Define $G = C_1$ and C_2, $H =$ not C_1 and not C_2. G and H are disjoint.
- (7): C_1 and C_2 are stochastically independent. Same for not C_1 and not C_2.

Calculating $P(X_2 = 2)$

$$\begin{aligned} P(X_2 = 2) &= P(X_1 = 0 \text{ and } X_2 = 2 \text{ or } X_1 = 1 \text{ and } X_2 = 2 \\ &\quad \text{or } X_1 = 2 \text{ and } X_2 = 2) \\ &= P(X_1 = 0 \text{ and } X_2 = 2) + P(X_1 = 1 \text{ and } X_2 = 2) \\ &\quad + P(X_1 = 2 \text{ and } X_2 = 2) \end{aligned}$$

- Since X_1 cannot be 0, $P(X_1 = 0 \text{ and } X_2 = 2)$ is 0.
- The second term, $P(X_1 = 1 \text{ and } X_2 = 2)$, we'll use (3).

$$P(X_1 = 1 \text{ and } X_2 = 2) = P(X_1 = 1)P(X_2 = 2 | X_1 = 1) \quad (9)$$

- $P(X_1 = 1)$: For the event in question to occur, either Node A would send and Node B wouldn't, or A would refrain from sending and Node B would send. Thus

$$P(X_1 = 1) = 2p(1 - p) = 0.48 \quad (10)$$

Calculating $P(X_2 = 2|X_1 = 1)$

Now, we need to find $P(X_2 = 2|X_1 = 1)$. This again involves breaking big events down into small ones. If $X_1 = 1$, then $X_2 = 2$ can occur only if *both* of the following occur:

- Event A: Whichever node was the one to successfully transmit during epoch 1—and we are given that there indeed was one, since $X_1 = 1$—now generates a new message.
- Event B: During epoch 2, no successful transmission occurs, i.e. either they both try to send or neither tries to send.

Recalling the definitions of p and q, we have that

$$P(X_2 = 2|X_1 = 1) = q\left[p^2 + (1-p)^2\right] = 0.41 \qquad (11)$$

Thus, $P(X_1 = 1 \text{ and } X_2 = 2) = 0.48 \times 0.41 = 0.20$.

Further Calculations and Interpretations

- Calculate $P(X_1 = 1|X_2 = 2)$ using conditional probability formula.
- Understanding the notebook view.
- Calculation of $P(X_1 = 2 \text{ or } X_2 = 2)$ using probability properties.
- Note on non-independence of events involving X_1 and X_2.

Example: Dice and Conditional Probability

- $P(B|A)$ and $P(A|B)$ are different quantities.
- $P(B|A)$ focuses on lines where event A occurs.
- $P(A|B)$ focuses on lines where event B occurs.
- Consider two dice rolls, resulting in random variables X and Y.
- Let $S = X + Y$, $T =$ number of even-dotted dice (i.e., number of dice having an even number of dots, 0, 1 or 2.).
- If $S = 12$, then $T = 2$ ($P(T = 2 \mid S = 12) = 1$).
- If $T = 2$, it doesn't imply $S = 12$ ($P(S = 12 \mid T = 2) < 1$).

ALOHA Experiment in the Notebook

Think of doing the ALOHA experiment many times:

- Run the network for two epochs, starting with both nodes active.
- Record outcomes in the notebook.
- Repeat for multiple iterations.

Notebook Outcomes

Line	$X_1 = 2$	$X_2 = 2$	$X_1 = 2$ and $X_2 = 2$	$X_2 = 2 \mid X_1 = 2$
1	Yes	No	No	No
2	No	No	No	NA
3	Yes	Yes	Yes	Yes
4	Yes	No	No	No
5	Yes	Yes	Yes	Yes
6	No	No	No	NA
7	No	Yes	No	NA

Table 3: Top of Notebook for Two-Epoch ALOHA Experiment

Observations from the Notebook

- Among the first seven lines, 4/7 have $X_1 = 2$, approaching 0.52 with many lines.
- Among the first seven lines, 3/7 have $X_2 = 2$, approaching 0.47 with many lines.
- Among the first seven lines, 2/7 have $X_1 = 2$ and $X_2 = 2$, approaching 0.27 with many lines.
- Among the first seven lines, 2/4 with non-NA $X_2 = 2 | X_1 = 2$ say Yes, approaching 0.52 with many lines.

A Note on Modeling

- Understand the ALOHA model and its parameters.
- Model properties captured:
  - Differences in network usage between A and B.
  - Adjusting model parameters to accommodate different behaviors.
- Modeling involves creative problem-solving.
- Build models by identifying important variables, using formulas, and reasoning through events.

Solution Strategies

- Naming important variables and events.
- Breaking down complex events into simpler ones.
- Adhering to conventions, using proper notation.
- Meticulous step-by-step approach for learning.
- Developing creative problem-solving skills.

Other Examples (see Chapter 1 of Matloff's book)

- Bus Ridership
- A Simple Board Game
- Document Classification
- Preferential Attachment Model
- Random Groups of Students
- Lottery Tickets
- Gaps between Numbers
- Probability of Getting Four Aces in a Bridge Hand

Conclusion I

Probability concepts provide powerful tools for understanding and analyzing real-world scenarios. The "notebook" perspective and strategic equations like (2) and (3) help simplify complex probability problems, making them more manageable.

- Probability concepts are foundational for diverse applications.
- The "notebook" view enhances intuition and simplifies problem-solving.
- Equations (2) and (3) offer a strategic approach to calculating probabilities.
- Probability analysis plays a critical role in various fields, from networking to finance and beyond.

Conclusion II

- Probability concepts applied to various examples.
- Creative problem-solving key in modeling.
- Understanding conditional probability, events, and modeling.
- Use proper notation, step-by-step approach for accurate solutions.

Bayes' Rule

(This section should not be confused with Section 8.7. The latter is highly controversial, while the material in this section is not controversial at all.)

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\text{not } A)P(B|\text{not } A)} \tag{12}$$

This is known as **Bayes' Theorem** or **Bayes' Rule**.

Combinatorics-Based Probability Computation

"And though the holes were rather small, they had to count them all"
- From the Beatles song, "A Day in the Life"

In some probability problems, all outcomes are **equally likely**. The probability computation is a matter of counting outcomes of interest and dividing by total possible outcomes. We'll discuss two examples here.

Five Card Probabilities

Suppose we deal a 5-card hand from a regular 52-card deck.
Which is larger: P(1 king) or P(2 hearts)?

Key Point: All possible hands are equally likely, so counting is the approach.

Probability of 1 king: $\frac{4 \cdot \binom{48}{4}}{\binom{52}{5}} = 0.299$

Probability of 2 hearts: $\frac{\binom{13}{2} \cdot \binom{39}{3}}{\binom{52}{5}} = 0.274$

The 1-king hand is slightly more likely.

Random Groups of Students

A class has 68 students: 48 CS majors, 20 others. Randomly assign 4 students to a group. Find the probability of exactly 2 CS majors in a group.

Probability: $\frac{\binom{48}{2} \cdot \binom{20}{2}}{\binom{68}{4}}$

Lottery Tickets

Twenty tickets are sold in a lottery, numbered 1 to 20. Five tickets are drawn for prizes. Find the probability of two even-numbered tickets winning.

Probability: $\frac{\binom{10}{2} \cdot \binom{10}{3}}{\binom{20}{5}}$

“Association Rules” in Data Mining

Data mining: extracting patterns from large databases. Market basket problem: finding patterns in sales transactions.

Association rules: $A, B \Rightarrow C, D, E$. How many possible rules with three or fewer antecedents out of 20 products?

Probability: $\frac{\sum_{k=1}^{3} \binom{20}{k} \cdot \binom{20-k}{1}}{\sum_{k=1}^{19} \binom{20}{k} \cdot \binom{20-k}{1}} = 0.0022$

Multinomial Coefficients

Question: Seating arrangements for 6 Democrats, 5 Republicans, and 2 Independents. How many? Using multinomial coefficients:

$\frac{13!}{6!5!2!}$

Multinomial Coefficients: $\frac{c!}{c_1!...c_r!}, \quad c_1 + ... + c_r = c$

Probability of Four Aces

A deck of 52 cards dealt to four players, 13 cards each. Probability that one player (Millie) gets all four aces:

Probability: $\frac{48!}{13!13!13!9!} / \frac{52!}{13!13!13!13!} = 0.00264$