Probability

Probability

Lecture objectives

- **Comprehend the different ways of assigning probability**
- **Understand and apply marginal, union, joint, and conditional probabilities**
- **Solve problems using the laws of probability including the laws of addition, multiplication and conditional probability**
- **Revise probabilities using Bayes' rule**

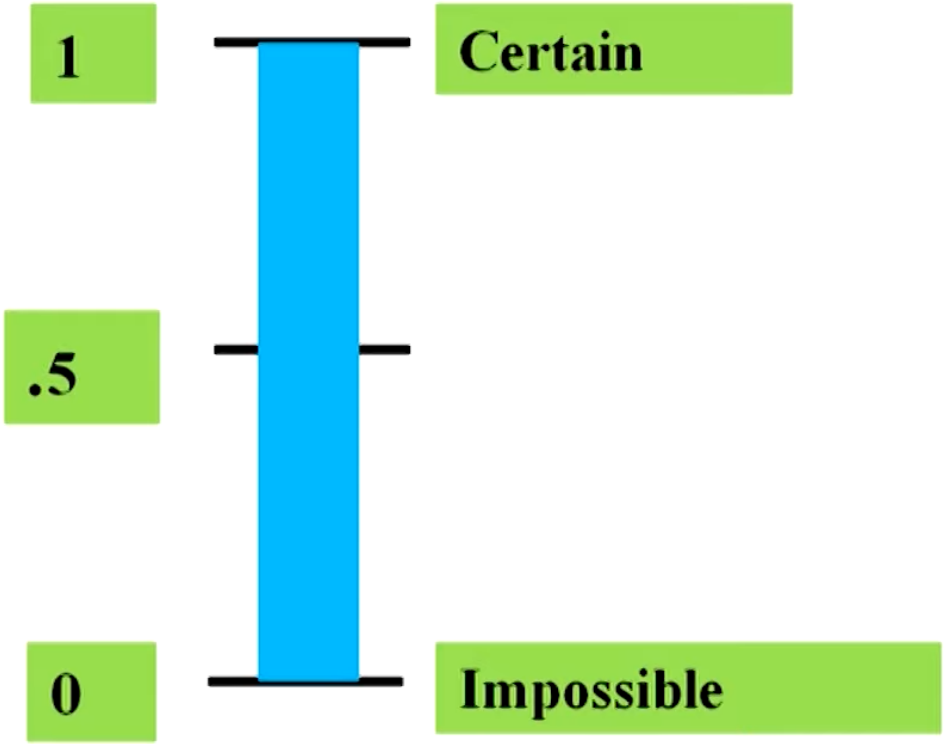
Probability

- Probability is the numerical measure of the likelihood that an event will occur.

- The probability of any event must be between 0 and 1, inclusively
  - $0 \leq P(A) \leq 1$ for any event A.

- The sum of the probabilities of all mutually exclusive and collectively exhaustive events is 1.
  - $P(A) + P(B) + P(C) = 1$
  - A, B, and C are mutually exclusive and collectively exhaustive

Range of Probability



Method of assigning probability

- Classical method of assigning probability (rules and laws)
- Relative frequency of occurrence (cumulated historical data)
- Subjective Probability (personal intuition or reasoning)

Classical probability

- Number of outcomes leading to the event divided by the total number of outcomes possible
- Each outcome is equally likely
- Determined *a priori* -- before performing the experiment
- Applicable to games of chance
- Objective -- everyone correctly using the method assigns an identical probability

Classical probability

$$P(E) = \frac{n_e}{N}$$

Where:

N = total number of outcomes

n_e = number of outcomes in E

Relative frequency probability

- Based on historical data
- Computed after performing the experiment
- Number of times an event occurred divided by the number of trials
- Objective -- everyone correctly using the method assigns an identical probability

Relative frequency probability

$$P(E) = \frac{n_e}{N}$$

Where :

N = total number of trials

n_e = number of outcomes producing E

Relative frequency probability

Relative Frequency

We calculate **relative frequency** using the following formula:

$$\text{relative frequency} = \frac{\text{no. of times an outcome happened}}{\text{total no. of all outcomes}}$$

Example: A coin is flipped 100 times, the coin lands on heads 48 times.

Calculate the relative frequency for heads.

$$\text{relative frequency} = \frac{48}{100} = 0.48$$

for landing on heads.

Relative frequency probability

Bias

- One of the key uses of relative frequency is in testing for bias. We say that an experiment is biased when the probability of a particular outcome is unfairly bigger or smaller than it should be.

- Example: A biased coin might be cleverly designed so that it's lands on tails more than 50% of the time. If we suspect a coin of being biased, then we test it by flipping it a lot of times and recording the results. Then, if the relative frequency of tails is noticeably more than 50%, we might suspect the coin of being biased

Subjective probability

- Comes from a person’s intuition or reasoning
- Subjective -- different individuals may (correctly) assign different numeric probabilities to the same event
- Degree of belief
- Useful for unique (single-trial) experiments
  - New product introduction
  - Initial public offering of common stock
  - Site selection decisions
  - Sporting events

Probability -Terminology

- Experiment
- Event
- Elementary Events
- Sample Space
- Unions and Intersections
- Mutually Exclusive Events
- Independent Events
- Collectively Exhaustive Events
- Complementary Events

Experiment, Trial, Elementary Event and Event

- Experiment: a process that produces outcomes
  - More than one possible outcome
  - Only one outcome per trial
- Trial: one repetition of the process
- Elementary Event: cannot be decomposed or broken down into other events
- Event: an outcome of an experiment
  - may be an elementary event, or
  - may be an aggregate of elementary events
  - usually represented by an uppercase letter, e.g., A, E1

An Example Experiment

- Experiment: randomly select, without replacement, two families from the residents of Tiny Town
- Elementary Event: the sample includes families A and C
- Event: each family in the sample has children in the household
- Event: the sample families own a total of four automobiles

Tiny Town Population		
Family	Children in Household	Number of Automobiles
A	Yes	3
B	Yes	2
C	No	1
D	Yes	2

Sample Space

- The set of all elementary events for an experiment
- Methods for describing a sample space
  - roster or listing
  - tree diagram
  - set builder notation
  - Venn diagram

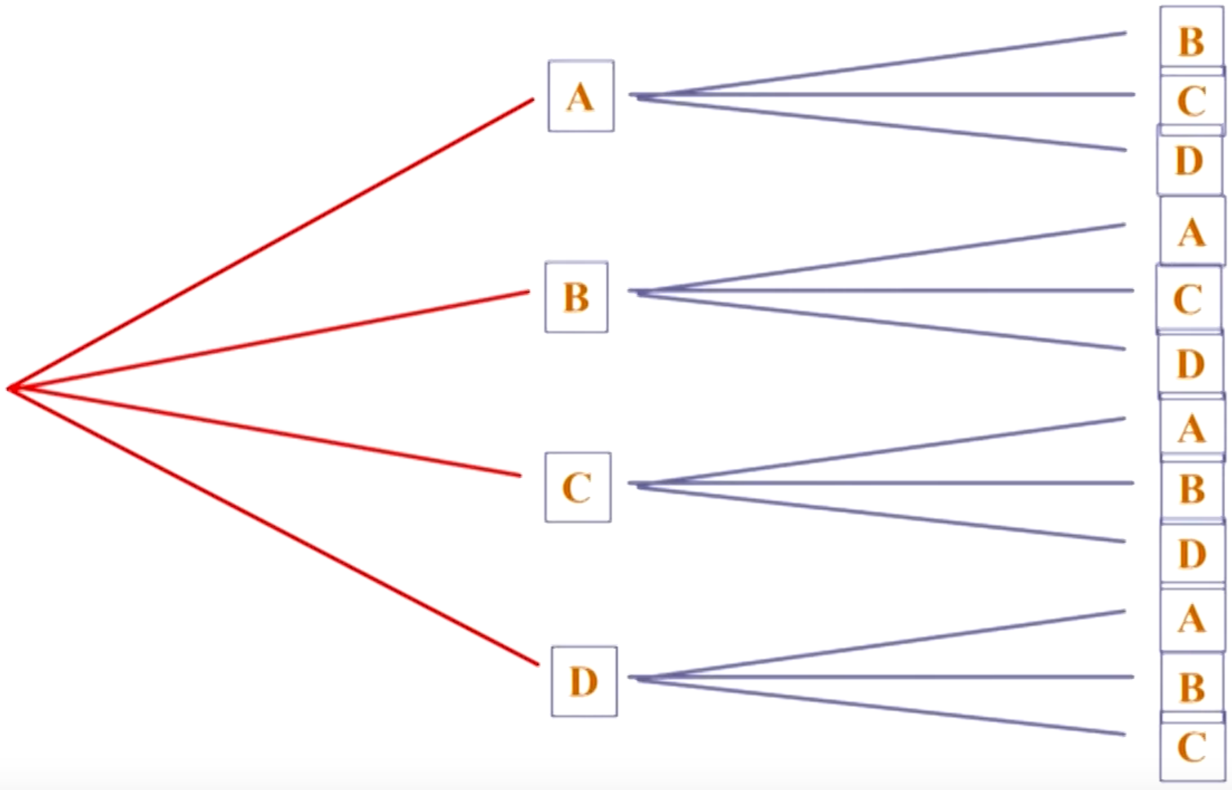
Sample Space–Roster Example

- Experiment: randomly select, without replacement, two families from the residents of Tiny Town
- Each ordered pair in the sample space is an elementary event, for example -- (D,C)

Family	Children in Household	Number of Automobiles
A	Yes	3
B	Yes	2
C	No	1
D	Yes	2

Listing of Sample Space
(A,B), (A,C), (A,D), (B,A), (B,C), (B,D), (C,A), (C,B), (C,D), (D,A), (D,B), (D,C)

Sample Space-Tree Diagram for Random sample of Two Families



Sample Space-Set Notation for Random sample of Two Families

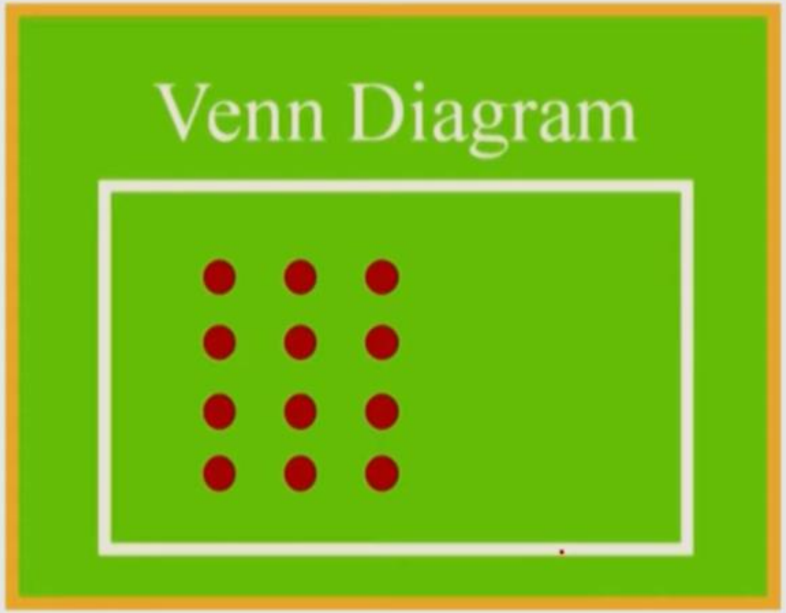
- S = {(x,y) | x is the family selected on the first draw, and y is the family selected on the second draw}
- Concise description of large sample spaces

Sample Space

- Useful for discussion of general principles and concepts

Listing of Sample Space

(A,B), (A,C), (A,D),
(B,A), (B,C), (B,D),
(C,A), (C,B), (C,D),
(D,A), (D,B), (D,C)

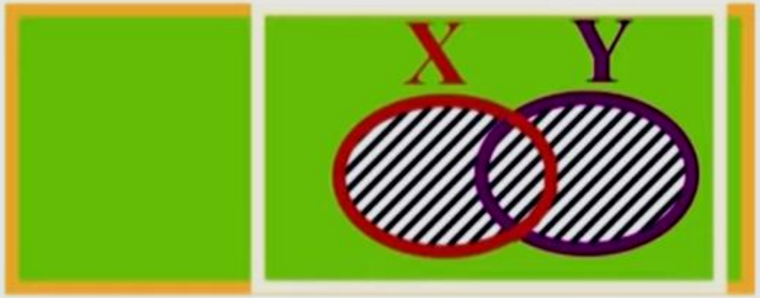


union of sets

- The union of two sets contains an instance of each element of the two sets.

$$X = \{1,4,7,9\}$$
$$Y = \{2,3,4,5,6\}$$
$$X \cup Y = \{1,2,3,4,5,6,7,9\}$$

$$C = \{IBM, DEC, Apple\}$$
$$F = \{Apple, Grape, Lime\}$$
$$C \cup F = \{IBM, DEC, Apple, Grape, Lime\}$$

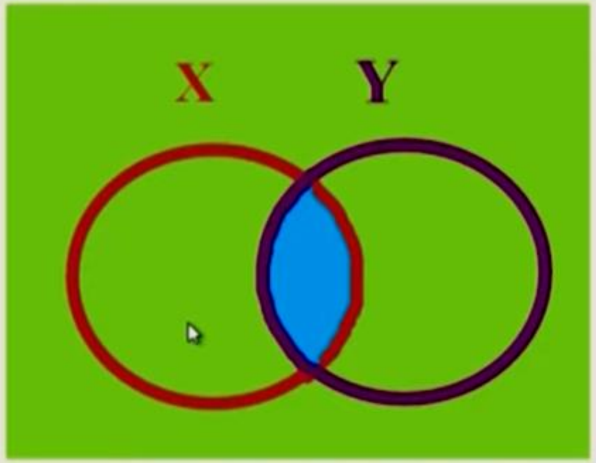


intersection of sets

- The intersection of two sets contains only those element common to the

$$X = \{1,4,7,9\}$$
$$Y = \{2,3,4,5,6\}$$
$$X \cap Y = \{4\}$$

$$C = \{IBM, DEC, Apple\}$$
$$F = \{Apple, Grape, Lime\}$$
$$C \cap F = \{Apple\}$$

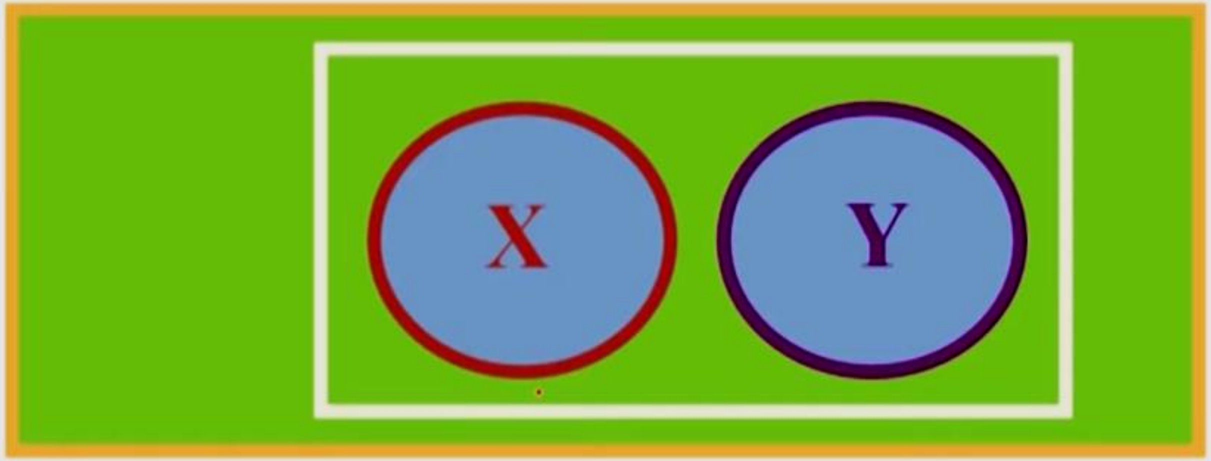


Mutually Exclusive Events

- Events with no common outcomes
- Occurrence of one event precludes the occurrence of the other event

$$C = \{IBM, DEC, Apple\}$$
$$F = \{Grape, Lime\}$$
$$C \cap F = \{\ \}$$

$$X = \{1, 7, 9\}$$
$$Y = \{2, 3, 4, 5, 6\}$$
$$X \cap Y = \{\ \}$$



$$P(X \cap Y) = 0$$

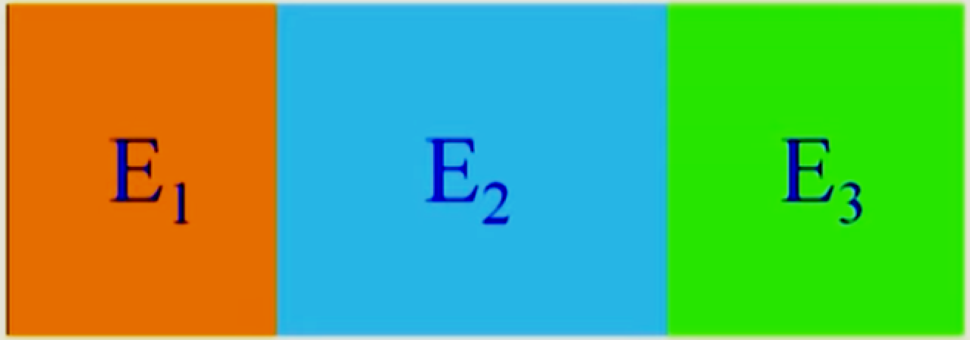
Independent Events

- Occurrence of one event does not affect the occurrence or nonoccurrence of the other event
- The conditional probability of X given Y is equal to the marginal probability of X.
- The conditional probability of Y given X is equal to the marginal probability of Y.

$$\boxed{P(X|Y) = P(X) \ and \ P(Y|X) = P(Y)}$$

Collectively exhaustive Events

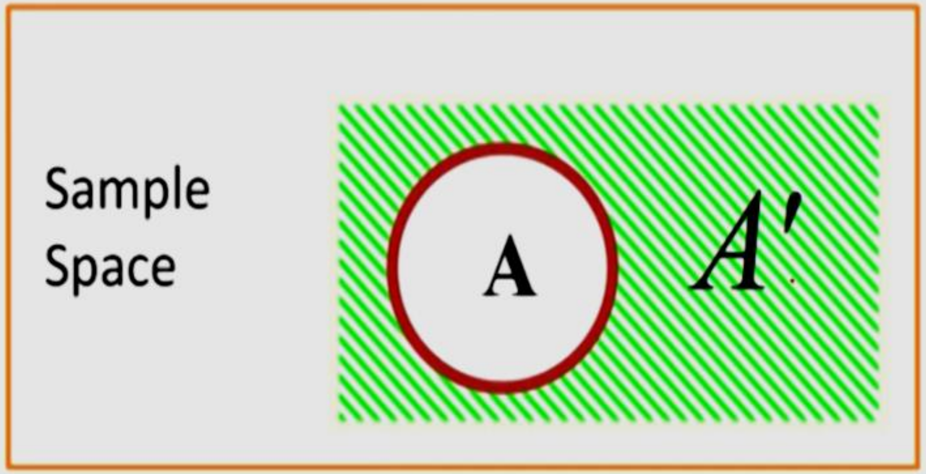
- Contains all elementary events for an experiment



Sample Space with three collectively exhaustive events

Complementary Events

- All elementary events not in the event 'A' are in its complementary event.



$$P(Sample\ Space) = 1$$

$$P(A') = 1 - P(A)$$

counting the possibilities

- mn Rule
- Sampling from a Population with Replacement
- Combinations: Sampling from a Population without Replacement

mn Rule

- If an operation can be done m ways and a second operation can be done n ways, then there are mn ways for the two operations to occur in order.
- This rule is easily extend to k stages, with a number of ways equal to

$n_1.n_2.n_3..n_k$

- Example: Toss two coins . The total umber of simple events is 2 x 2 =4

Sampling from a Population with Replacement

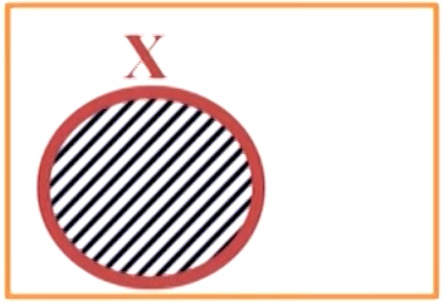
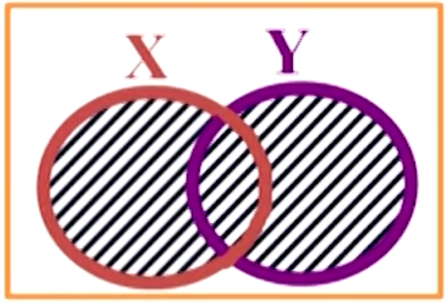
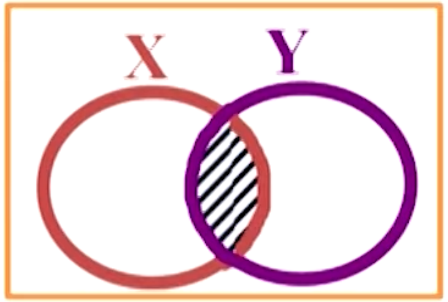
- A tray contains 1,000 individual tax returns. If 3 returns are randomly selected with replacement from the tray, how many possible samples are there?
- $(N)^n = (1{,}000)^3 = 1{,}000{,}000{,}000$

Combinations

- A tray contains 1,000 individual tax returns. If 3 returns are randomly selected without replacement from the tray, how many possible samples are there?

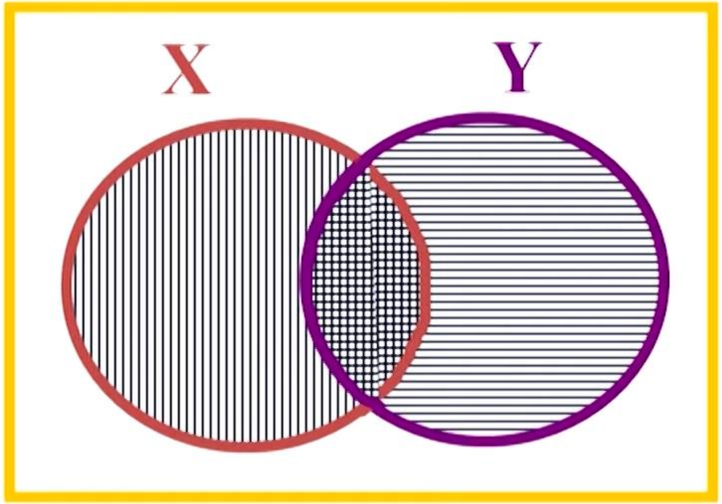
$$\boxed{\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{1000!}{3!(1000-3)!} = 166,167,000}$$

Four Types of Probability

Marginal	Union	Joint	Conditional
$P(X)$	$P(X \cup Y)$	$P(X \cap Y)$	$P(X \mid Y)$
The probability of **X** occurring	The probability of **X or Y** occurring	The probability of **X and Y** occurring	The probability of **X occurring given that Y** has occurred
			

General Law of Addition

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$



Design for improving the productivity

Problem

- A company conducted a survey for the American Society of Interior Designers in which workers were asked which changes in office design would increase productivity.
- Respondents were allowed to answer more than one type of design change.

Reducing noise would increase productivity	70 %
More storage space would increase productivity	67 %

Problem

- If one of the survey respondents was randomly selected and asked what office design changes would increase worker productivity,
  - what is the probability that this person would select reducing noise *or* more storage space?

EduBridge

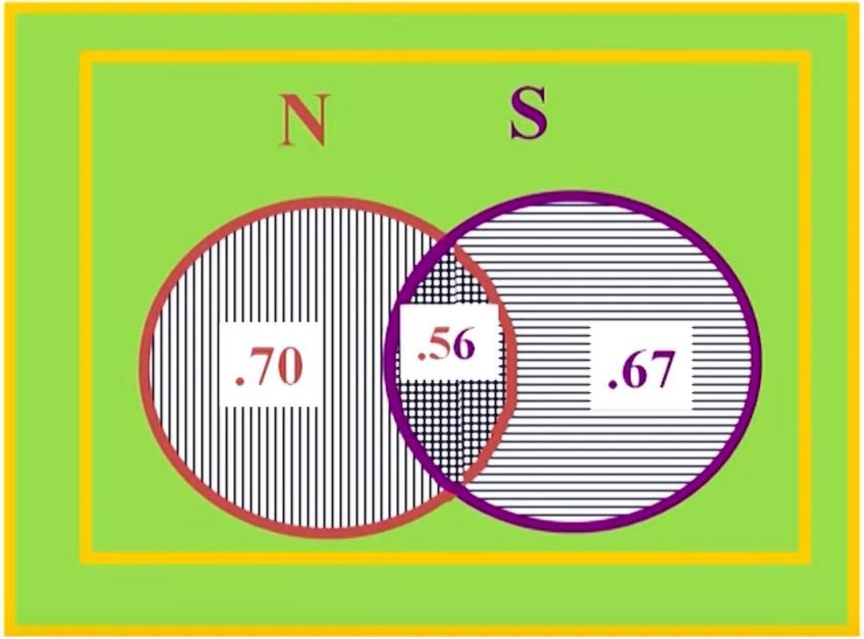
Solution

- Let *N* represent the event "reducing noise."
- Let *S* represent the event "more storage/ filing space."
- The probability of a person responding with *N or S* can be symbolized statistically as a union probability by using the law of addition.

$$P(N \cup S)$$

Solution- General Law of Addition

$$P(N \cup S) = P(N) + P(S) - P(N \cap S)$$



$$P(N) = .70$$
$$P(S) = .67$$
$$P(N \cap S) = .56$$
$$P(N \cup S) = .70 + .67 - .56$$
$$= 0.81$$

Office Design Problem
Probability Matrix

		Increase Storage Space		
		Yes	No	Total
Noise Reduction	Yes	.56	.14	.70
	No	.11	.19	.30
	Total	.67	.33	1.00

Joint Probability Using a Contingency Table

Event	Event B_1	Event B_2	Total
A_1	$P(A_1 \text{ and } B_1)$	$P(A_1 \text{ and } B_2)$	$P(A_1)$
A_2	$P(A_2 \text{ and } B_1)$	$P(A_2 \text{ and } B_2)$	$P(A_2)$
Total	$P(B_1)$	$P(B_2)$	1

Joint Probabilities

Marginal (Simple) Probabilities

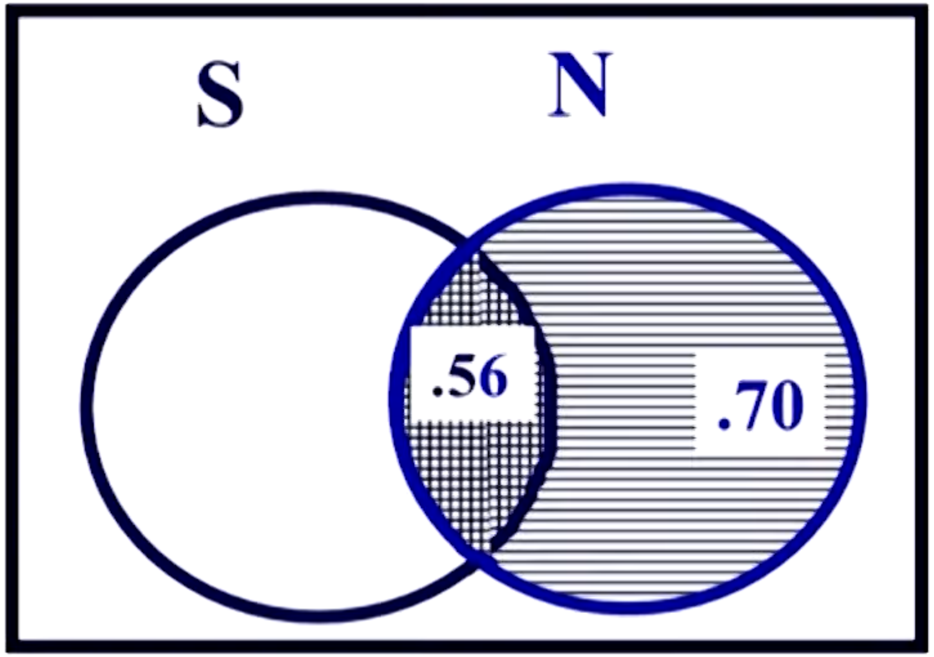
S

Office Design Problem - Probability Matrix

		Increase Storage Space		
		Yes	No	Total
Noise Reduction	Yes	.56	.14	.70
	No	.11	.19	.30
	Total	.67	.33	1.00

$$
\begin{aligned}
P(N \cup S) &= P(N) + P(S) - P(N \cap S) \\
&= .70 + .67 - .56 \\
&= .81
\end{aligned}
$$

Law of Conditional Probability



$$P(N) = .70$$

$$P(N \cap S) = .56$$

$$P(S|N) = \frac{P(N \cap S)}{P(N)}$$

$$= \frac{.56}{.70}$$

$$= .80$$

Solution

Office Design Problem

		Increase Storage Space		
		Yes	No	Total
Noise Reduction	Yes	.56	.14	.70
	No	.11	.19	.30
	Total	.67	.33	1.00

$$P(\overline{N} \mid S) = \frac{P(\overline{N} \cap S)}{P(S)} = \frac{.11}{.67}$$

$$= .164$$

Problem

- A company data reveal that 155 employees worked one of four types of positions.
- Shown here again is the raw values matrix (also called a contingency table) with the frequency counts for each category and for subtotals and totals containing a breakdown of these employees by type of position and by sex.

Contingency Table

COMPANY HUMAN RESOURCE DATA

Type of Position		Sex: Male	Sex: Female	
	Managerial	8	3	11
	Professional	31	13	44
	Technical	52	17	69
	Clerical	9	22	31
		100	55	155

- If an employee of the company is selected randomly, what is the probability that the employee is female or a professional worker?

$$P(F \cup P) = P(F) + P(P) - P(F \cap P)$$

$$P(F \cup P) = .355 + .284 - .084 = .555.$$

Law of Multiplication

$$\boxed{P(X \cap Y) = P(X) \cdot P(Y|X) = P(Y) \cdot P(X|Y)}$$

Problem

- A company has 140 employees, of which 30 are supervisors.
- Eighty of the employees are married, and 20% of the married employees are supervisors.
- If a company employee is randomly selected, what is the probability that the employee is married and is a supervisor?

		Married		
		Y	N	Sub total
Supervisor	Y	0.1143		30
	N			110
	Sub total	80	60	140

$$P(M) = \frac{80}{140} = 0.5714$$

$$P(S|M) = 0.20$$

$$P(M \cap S) = P(M) \cdot P(S|M)$$

$$= (0.5714)(0.20) = 0.1143$$

Solution

Probability Matrix of Employees

Supervisor	Married: Yes	Married: No	Total
Yes	.1143	.1000	.2143
No	.4571	.3286	.7857
Total	.5714	.4286	1.00

$$P(\bar{S}) = 1 - P(S)$$
$$= 1 - 0.2143 = 0.7857$$
$$P(\overline{M} \cap \bar{S}) = P(\bar{S}) - P(M \cap \bar{S})$$
$$= 0.7857 - 0.4571 = 0.3286$$

$$P(M \cap \bar{S}) = P(M) - P(M \cap S)$$
$$= 0.5714 - 0.1143 = 0.4571$$
$$P(\overline{M} \cap S) = P(S) - P(M \cap S)$$
$$= 0.2143 - 0.1143 = 0.1000$$
$$P(\overline{M}) = 1 - P(M)$$
$$= 1 - 0.5714 = 0.4286$$

Special Law of Multiplication for independent Events

- General Law

$$P(X \cap Y) = P(X) \cdot P(Y \mid X) = P(Y) \cdot P(X \mid Y)$$

- Special Law

If events X and Y are independent,

$P(X) = P(X \mid Y)$, and $P(Y) = P(Y \mid X)$.

Consequently,

$P(X \cap Y) = P(X) \cdot P(Y)$

Law of Conditional Probability

- The conditional probability of X given Y is the joint probability of X and Y divided by the marginal probability of Y.

$$\boxed{P(X|Y) = \frac{P(X \cap Y)}{P(Y)} = \frac{P(Y|X) \cdot P(X)}{P(Y)}}$$

Conditional Probability

- A conditional probability is the probability of one event, given that another event has occurred:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$ → The conditional probability of A given that B has occurred

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$ → The conditional probability of B given that A has occurred

Where P(A and B) = joint probability of A and B

Computing Conditional Probability

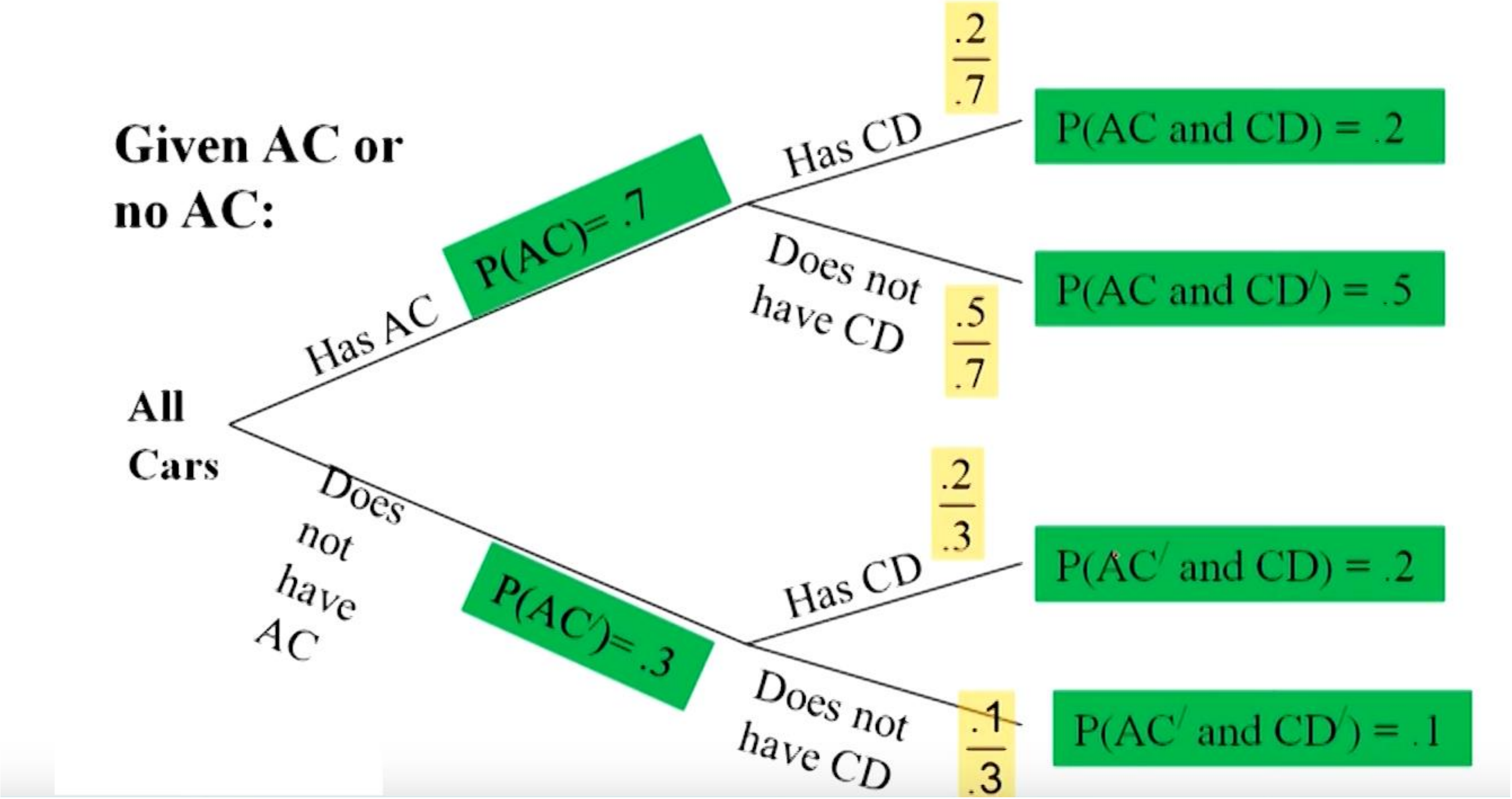
- Of the cars on a used car lot, 70% have air conditioning (AC) and 40% have a CD player (CD). 20% of the cars have both.
- What is the probability that a car has a CD player, given that it has AC ?
- We want to find P(CD | AC).

	CD	No CD	Total
AC	0.2	0.5	0.7
No AC	0.2	0.1	0.3
Total	0.4	0.6	1.0

$$P(CD \mid AC) = \frac{P(CD \text{ and } AC)}{P(AC)} = \frac{.2}{.7} = .2857$$

Given AC, we only consider the top row (70% of the cars). Of these, 20% have a CD player. 20% of 70% is about 28.57%.

Computing Conditional Probability- Decision Tree



Independent Events

- If X and Y are independent events, the occurrence of Y does not affect the probability of X occurring.
- If X and Y are independent events, the occurrence of X does not affect the probability of Y occurring.

If X and Y are independent events

$$P(X|Y) = P(X), \text{ and}$$

$$P(Y|X) = P(Y).$$

Statistical Independence

- Two events are **independent** if and only if:

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

- Events A and B are independent when the probability of one event is not affected by the other event

Independent Event - Demo

	Geographic Location				
	Northeast D	Southeast E	Midwest F	West G	
Finance A	.12	.05	.04	.07	.28
Manufacturing B	.15	.03	.11	.06	.35
Communications C	.14	.09	.06	.08	.37
	.41	.17	.21	.21	1.00

Test the matrix for the 200 executive responses to determine whether industry type is independent of geographic location.

$$P(A|G) = \frac{P(A \cap G)}{P(G)} = \frac{0.07}{0.21} = 0.33 \qquad P(A) = 0.28$$

$$P(A|G) = 0.33 \neq P(A) = 0.28$$

Independent Event - Demo

	D	E	
A	8	12	20
B	20	30	50
C	6	9	15
	34	51	85

$$P(A|D) = \frac{8}{34} = .2353$$

$$P(A) = \frac{20}{85} = .2353$$

$$P(A|D) = P(A) = 0.2353$$

Revision of Original Probabilities- Bayes Rule

- An extension to the conditional law of probabilities
- Enables revision of original probabilities with new information

$$P(X_i|Y) = \frac{P(Y|X_i)P(X_i)}{P(Y|X_1)P(X_1) + P(Y|X_2)P(X_2) + \cdots P(Y|X_n)P(X_n)}$$

Problem

- Machines A, B, and C all produce the same two parts, X and Y. Of all the parts produced, machine A produces 60%, machine B produces 30%, and machine C produces 10%. In addition
  - 40% of the parts made by machine A are part X.
  - 50% of the parts made by machine B are part X.
  - 70% of the parts made by machine C are part X.
- A part produced by this company is randomly sampled and is determined to be an X part.
- With the knowledge that it is an X part, revise the probabilities that the part came from machine A, B, or C.

Solution

Event	Prior $P(E_i)$	Conditional $P(X \mid E_i)$	Joint $P(X \cap E_i)$	Posterior
A	.60	.40	$(.60)(.40) = .24$	$\frac{.24}{.46} = .52$
B	.30	.50	.15	$\frac{.15}{.46} = .33$
C	.10	.70	.07	$\frac{.07}{.46} = .15$
			$P(X) = .46$	

Problem

- A particular type of printer ribbon is produced by only two companies, Alamo Ribbon Company and South Jersey Products.
- Suppose Alamo produces 65% of the ribbons and that South Jersey produces 35%.
- Eight percent of the ribbons produced by Alamo are defective and 12% of the South Jersey ribbons are defective
- A customer purchases a new ribbon. What is the probability that Alamo produced the ribbon? What is the probability that South Jersey produced the ribbon?

Solution

$$P(Alamo) = 0.65$$

$$P(SouthJersey) = 0.35$$

$$P(d|Alamo) = 0.08$$

$$P(d|SouthJersey) = 0.12$$

$$P(Alamo|d) = \frac{P(d|Alamo) \cdot P(Alamo)}{P(d|Alamo) \cdot P(Alamo) + P(d|SouthJersey) \cdot P(SouthJersey)}$$

$$= \frac{(0.08)(0.65)}{(0.08)(0.65) + (0.12)(0.35)} = 0.553$$

$$P(SouthJersey|d) = \frac{P(d|SouthJersey) \cdot P(SouthJersey)}{P(d|Alamo) \cdot P(Alamo) + P(d|SouthJersey) \cdot P(SouthJersey)}$$

$$= \frac{(0.12)(0.35)}{(0.08)(0.65) + (0.12)(0.35)} = 0.447$$

Solution

Event	Prior Probability $P(E_i)$	Conditional Probability $P(d\|E_i)$	Joint Probability $P(E_i \cap d)$	Revised Probability $P(E_i\|d)$
Alamo	0.65	0.08	0.052	$\frac{0.052}{0.094}$ =0.553
South Jersey	0.35	0.12	$\underline{0.042}$ 0.094	$\frac{0.042}{0.094}$ =0.447

Solution


Defective
0.08
0.052
Alamo
0.65
Acceptable
0.92
0.094
Defective
0.12
0.042
South
Jersey
0.35
Acceptable
0.88