CONTINGENCY (CROSS-TABULATION) TABLES

- Presents counts of two or more variables

	A_1	A_2	Total
B_1	a	b	a+b
B_2	c	d	c+d
Total	a+c	b+d	n = a+b+c+d

Joint, Marginal, and Conditional Probability

- We study methods to determine probabilities of events that result from combining other events in various ways.
- There are several types of combinations and relationships between events:
  - Intersection of events
  - Union of events
  - Dependent and independent events
  - Complement event

Joint, Marginal, and Conditional Probability

- Joint probability is the probability that two events will occur simultaneously.
- Marginal probability is the probability of the occurrence of the single event.

The joint prob. of A_2 and B_1

	A_1	A_2	Total
B_1	a/n	b/n	(a+b)/n
B_2	c/n	d/n	(c+d)/n
Total	(a+c)/n	(b+d)/n	1

The marginal probability of A_1.

Intersection

- Example 1
  - A potential investor examined the relationship between the performance of mutual funds and the school the fund manager earned his/her MBA.
  - The following table describes the joint probabilities.

	Mutual fund outperform the market	Mutual fund doesn’t outperform the market
Top 20 MBA program	.11	.29
Not top 20 MBA program	.06	.54

Intersection

- Example 1 – continued
  - The joint probability of [mutual fund outperform…] **and** […from a top 20 …] = .11

P(A_1 and B_1)

  - The joint probability of [mutual fund outperform…] **and** […not from a top 20 …] = .06

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)
Top 20 MBA program (A_1)	.11	.29
Not top 20 MBA program (A_2)	.06	.54

Intersection

- Example 1 – continued
  - The joint probability of [mutual fund outperform…] **and** […from a top 20 …] = .11

    $P(A_1 \text{ and } B_1)$
  - The joint probability of [mutual fund outperform…] **and** […not from a top 20 …] = .06

    $P(A_2 \text{ and } B_1)$

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)
Top 20 MBA program (A_1)	.11	.29
Not top 20 MBA program (A_2)	.06	.54

Marginal Probability

- These probabilities are computed by adding across rows and down columns

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)	Marginal Prob. $P(A_i)$
Top 20 MBA program (A_1)	$P(A_1 \text{ and } B_1)$	$+\ P(A_1 \text{ and } B_2)$	$= P(A_1)$
Not top 20 MBA program (A_2)	$P(A_2 \text{ and } B_1)$	$+\ P(A_2 \text{ and } B_2)$	$= P(A_2)$
Marginal Probability $P(B_j)$			

Marginal Probability

- These probabilities are computed by adding across rows and down columns

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)	Marginal Prob. $P(A_i)$
Top 20 MBA program (A_1)	.11 +	.29 =	.40
Not top 20 MBA program (A_2)	.06 +	.54 =	.60
Marginal Probability $P(B_j)$			

Marginal Probability

- These probabilities are computed by adding across rows and down columns

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)	Marginal Prob. $P(A_i)$
Top 20 MBA program (A_1)	$P(A_1 \text{ and } \mathbf{B_1})$ + $P(A_2 \text{ and } \mathbf{B_1})$ = $P(\mathbf{B_1})$	$P(A_1 \text{ and } \mathbf{B_2})$ + $P(A_2 \text{ and } \mathbf{B_2})$ = $P(\mathbf{B_2})$	.40
Not top 20 MBA program (A_2)			.60
Marginal Probability $P(B_j)$			

Marginal Probability

- These probabilities are computed by adding across rows and down columns

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)	Marginal Prob. $P(A_i)$
Top 20 MBA program (A_1)	.11 +	.29 +	.40
Not top 20 MBA program (A_2)	.06	.54	.60
Marginal Probability $P(B_j)$	.17	.83	

Conditional Probability

- Example 2 (Example 1 – continued)
  - Find the conditional probability that a randomly selected fund is managed by a "Top 20 MBA Program graduate", given that it did not outperform the market.
- Solution

$$P(A_1|B_2) = \frac{P(A_1 \text{ and } B_2)}{P(B_2)} = \frac{.29}{.83} = 0.3949$$

CONDITIONAL PROBABILITY

- The probability that event A will occur given that or on the condition that, event B has already occurred. It is denoted by P(A|B).

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

Conditional Probability

- Example 2
  - Find the conditional probability that a randomly selected fund is managed by a “Top 20 MBA Program graduate”, given that $\boxed{\text{it did not outperform the mar}}$ket.
- Solution

New information reduces the relevant sample space to the 83% of event B_2.

$P(A_1|B_2) = \dfrac{P(A_1 \text{ and } B_2)}{P(B_2)}$

$= .29/.83 = .39$

	Mutual fund outperforms the market (B_1)	Mutual fund doesn’t outperform the market (B_2)	Marginal Prob. $P(A_i)$
Top 20 MBA program (A_1)	.11	.29	.40
Not top 20 MBA program (A_2)	.06	.54	.60
Marginal Probability $P(B_j)$	.17	.83	

Conditional Probability

- Before the new information becomes available we have

$$P(A_1) = 0.40$$

- After the new information becomes available $P(A_1)$ changes to

$$P(A_1 \text{ given } B_2) = .3494$$

- Since the the occurrence of B_2 has changed the probability of A_1, the two event are related and are called "**dependent events**".

EXAMPLE 3

The director of an insurance company's computing center estimates that the company's computer has a 20% chance of catching a computer virus. However, she feels that there is only a 6% chance of the computer's catching a virus that will completely disable its operating system. If the company's computer should catch a virus, what is the probability that the operating system will be completely disabled?

EXAMPLE 4

- Of a company's employees, 30% are women and 6% are married women. Suppose an employee is selected at random. If the employee selected is a woman, what is the probability that she is married?

Independence

- Independent events
  - Two events A and B are said to be **independent** if

$$P(A|B) = P(A)$$

or

$$P(B|A) = P(B)$$

  - That is, the probability of one event is not affected by the occurrence of the other event.

Dependent and independent events

- Example 5 (Example 1 – continued)
  - We have already seen the dependency between A_1 and B_2.
  - Let us check A_2 and B_2.
    - $P(B_2) = .83$
    - $P(B_2|A_2)=P(B_2 \text{ and } A_2)/P(A_2) = .54/.60 = .90$
  - Conclusion: A_2 and B_2 are dependent.

Union

Example 6 (Example 1 – continued) Calculating P(A or B))

– Determine the probability that a randomly selected fund outperforms the market or the manager graduated from a top 20 MBA Program.

Union

- Solution

Comment: $P(A_1 \text{ or } B_1) = 1 - P(A_2 \text{ and } B_2) = 1 - .46 = .54$	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)
Top 20 MBA program (A_1)	.11	.29
Not top 20 MBA program (A_2)	.06	.54

A_1 or B_1 occurs whenever either:
A_1 and B_1 occurs,
A_1 and B_2 occurs,
A_2 and B_1 occurs.

$P(A_1 \text{ or } B_1) = P(A_1 \text{ and } B_1) + P(A_1 \text{ and } B_2) + P(A_2 \text{ and } B_1) =$
$.11 + .29 + .06 = .46$

EXAMPLE 7

- There are three approaches to determining the probability that an outcome will occur: classical, relative frequency, and subjective. Which is most appropriate in determining the probability of the following outcomes?
- The unemployment rate will rise next month.
- Five tosses of a coin will result in exactly two heads.
- An American will win the French Open Tennis Tournament in the year 2000.
- A randomly selected woman will suffer a breast cancer during the coming year.

EXAMPLE 8

- Abby, Brenda, and Cameron; three candidates for the presidency of a college's student body, are to address a student forum. The forum's organizer is to select the order in which the candidates will give their speeches, and must do so in such a way that each possible order is equally likely to be selected.

A) What is the random experiment?

B) List the outcomes in the sample space.

C) Assign probabilities to the outcomes.

D) What is the probability that Cameron will speak first?

E) What is the probability that one of the women will speak first?

F) What is the probability that Abby will speak before Cameron does?

EXAMPLE 9

- Suppose A and B are two independent events for which $P(A) = 0.20$ and $P(B) = 0.60$.
- Find $P(A/B)$.
- Find $P(B/A)$.
- Find $P(A \text{ and } B)$.
- Find $P(A \text{ or } B)$.

EXAMPLE 10

- A Ph.D. graduate has applied for a job with two universities: *A* and *B*. The graduate feels that she has a 60% chance of receiving an offer from university *A* and a 50% chance of receiving an offer from university *B*. If she receives an offer from university *B*, she believes that she has an 80% chance of receiving an offer from university *A*.

a) What is the probability that both universities will make her an offer?

b) What is the probability that at least one university will make her an offer?

c) If she receives an offer from university *B*, what is the probability that she will not receive an offer from university *A*?

EXAMPLE 11

- Suppose $P(A) = 0.50$, $P(B) = 0.40$, and $P(B/A) = 0.30$.

a) Find $P(A \text{ and } B)$.

b) Find $P(A \text{ or } B)$.

c) Find $P(A/B)$.

EXAMPLE 12

- A statistics professor classifies his students according to their grade point average (GPA) and their gender. The accompanying table gives the proportion of students falling into the various categories. One student is selected at random.

Gender	**Under 2.0**	**2.0 – 3.0**	**Over 3.0**
Male	0.05	0.25	0.10
Female	0.10	0.30	0.20

- If the student selected is female, what is the probability that her GPA is between 2.0 and 3.0?
- If the GPA of the student selected is over 3.0, what is the probability that the student is male?
- What is the probability that the student selected is female or has a GPA under 2.0 or both?
- Is GPA independent of gender? Explain using probabilities.

Probability Rules and Trees

- We present more methods to determine the probability of the intersection and the union of two events.
- Three rules assist us in determining the probability of complex events from the probability of simpler events.

Multiplication Rule

- For any two events A and B

$$P(A \text{ and } B) = P(A)P(B|A)$$
$$= P(B)P(A|B)$$

- When A and B are independent

$$P(A \text{ and } B) = P(A)P(B)$$

Multiplication Rule

- Example 13

  What is the probability that two female students will be selected at random to participate in a certain research project, from a class of seven males and three female students?

- Solution
  - Define the events:
    A – the first student selected is a female
    B – the second student selected is a female
  - P(A and B) = P(A)P(B|A) = (3/10)(2/9) = 6/90 = .067

Multiplication Rule

- Example 14

  What is the probability that a female student will be selected at random from a class of seven males and three female students, in each of the next two class meetings?

- Solution
  - Define the events:

    A – the first student selected is a female

    B – the second student selected is a female
  - P(A and B) = P(A)P(B) = (3/10)(3/10) = 9/100 = .09

Addition Rule

For any two events A and B

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

P(A) =6/13

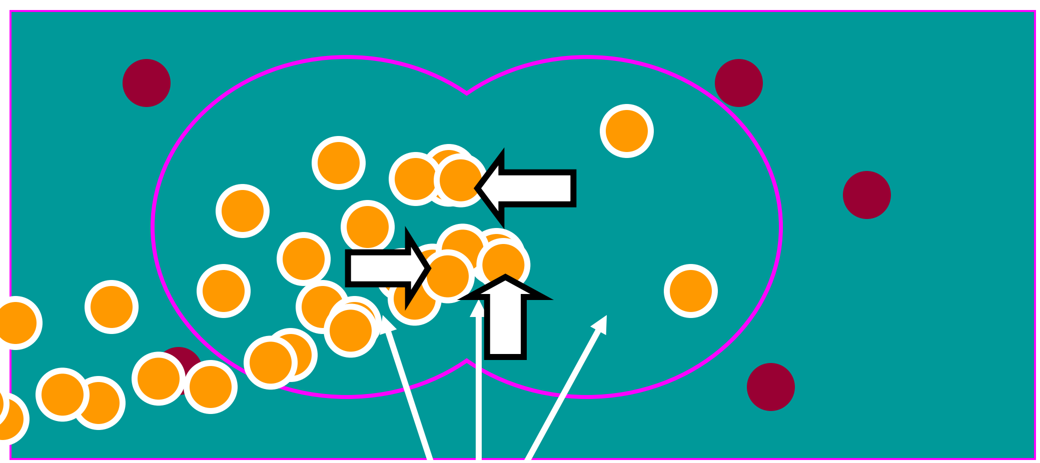
+

P(B) =5/13

—

P(A and B) = 3/13

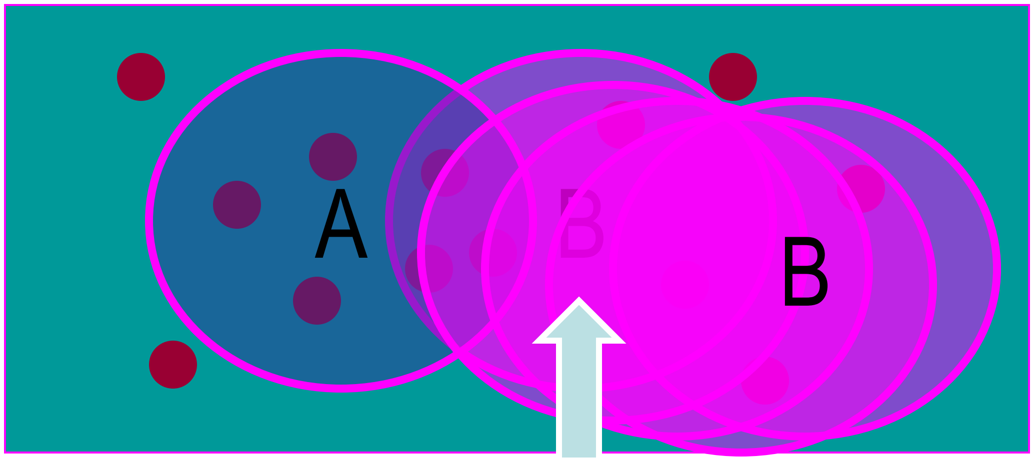
P(A or B) = 8/13



Addition Rule

When A and B are mutually exclusive,

$$P(A \text{ or } B) = P(A) + P(B)$$



P(A and B) = 0

Addition Rule

- Example 15
  - The circulation departments of two newspapers in a large city report that 22% of the city's households subscribe to the Sun, 35% subscribe to the Post, and 6% subscribe to both.
  - What proportion of the city's household subscribe to either newspaper?

Addition Rule

- Solution
  - Define the following events:
    - A = the household subscribe to the Sun
    - B = the household subscribe to the Post
  - Calculate the probability

    P(A or B) = P(A) + P(B) – P(A and B) = .22+.35 - .06 = .51

EXAMPLE

- A computer manufacturer inspected memory chips 100% before they enter assembly operations. Let

D: Defective chip

D^*: Non-defective chip

A: A chip approved for assembly by inspector

A^*: A chip not approved for assembly by inspector

From past experience, it is known that P(D)=0.10. Also, it is known that the probability of an inspector passing a chip given that it is defective is 0.005, while the corresponding probability, given that the chip is non-defective is 0.999.

EXAMPLE (contd.)

a) Find the joint probability that a chip is defective and is approved for assembly.

b) Find the probability that a chip is acceptable and is approved for assembly.

c) Find the probability that a chip is approved by assembly.

EXAMPLE

- The accompanying contingency table gives frequencies for a classification of the equipment used in a manufacturing plant.

	Equipment use			
Working status	Low	Moderate	High	**Total**
In working order	10	18	12	40
Under repair	2	6	8	16
Total	12	24	20	56

a) Find the probability that a randomly selected piece of equipment is a high-use item given that it is in working order.

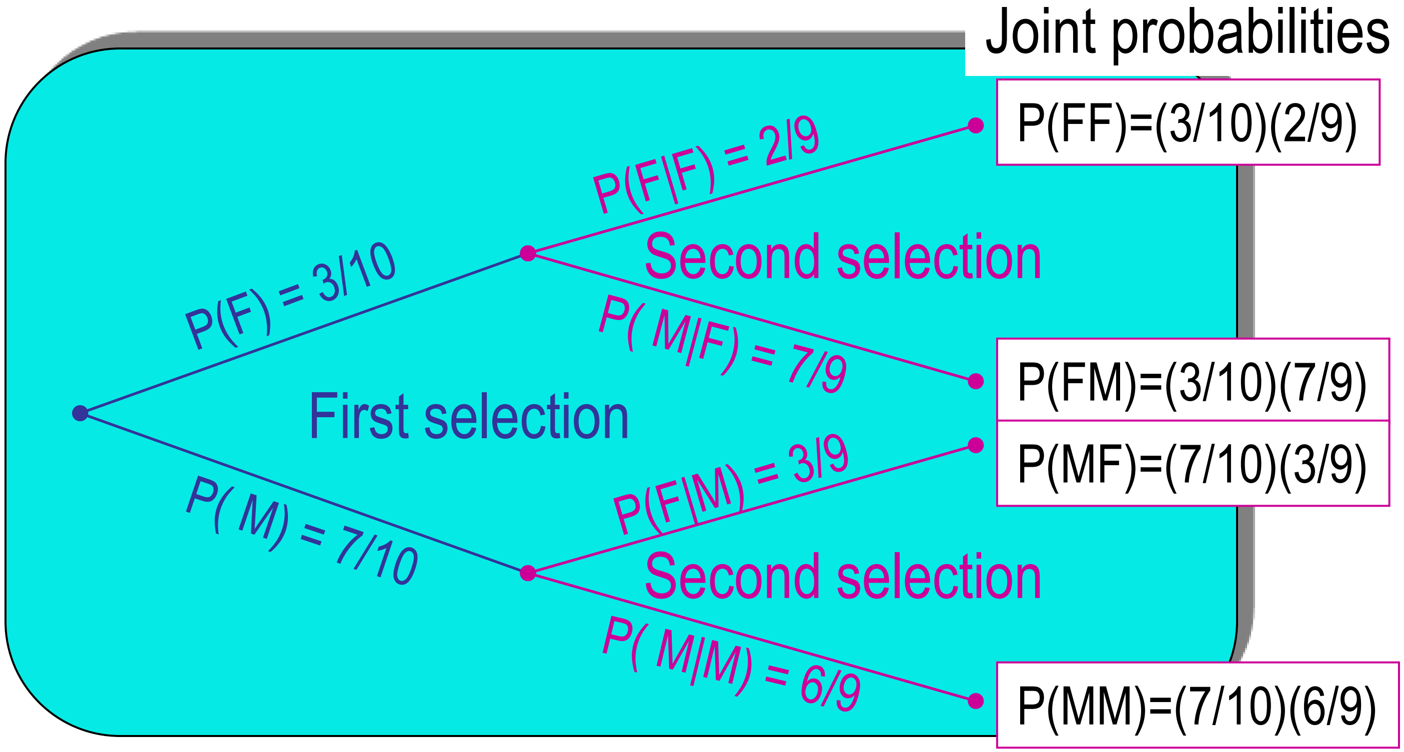
b) Find the probability that a randomly selected piece of equipment is under repair given that it is a moderate use item.

Probability Trees

- This is a useful device to calculate probabilities when using the probability rules.

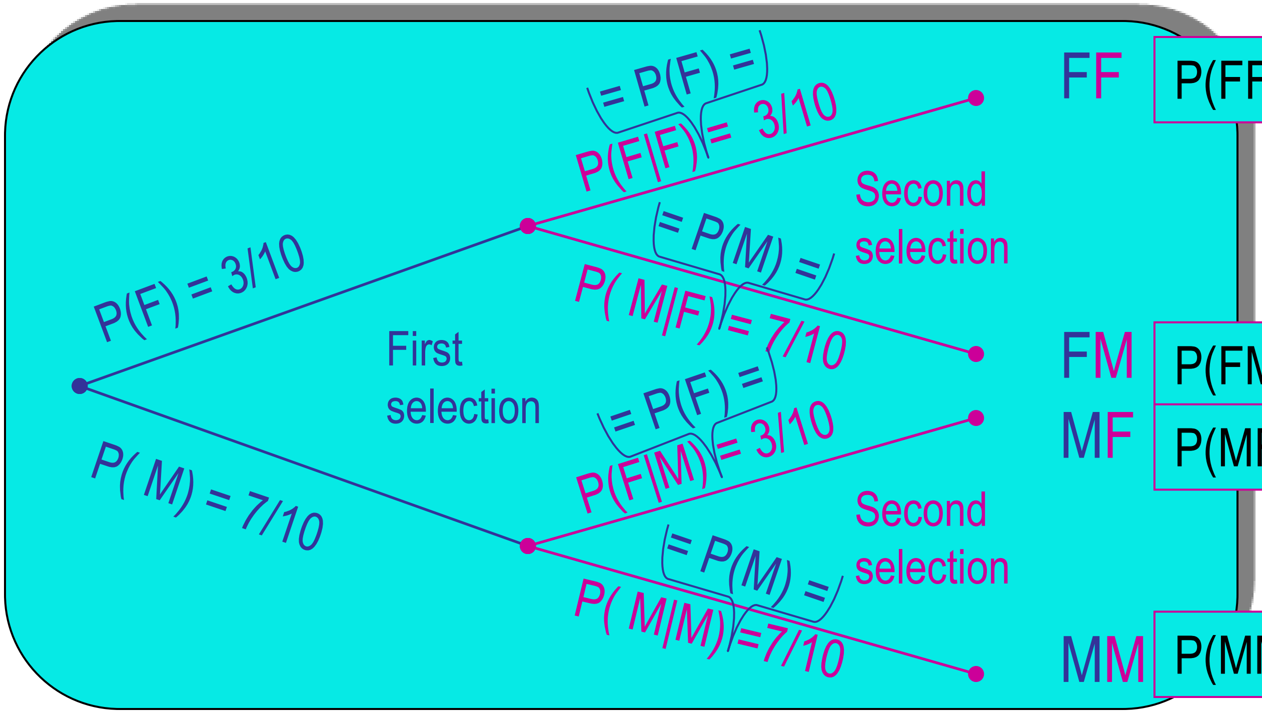
Probability Trees

- **Example 14 revisited (dependent events).**
  - Find the probability of selecting two female students (without replacement), if there are 3 female students in a class of 10.



Probability Trees

- **Example 15 – revisited (independent events)**
  - Find the probability of selecting two female students (with replacement), if there are 3 female students in a class of 10.



P(FF)=(3/10)(3/10)

P(FM)=(3/10)(7/10)

P(MF)=(7/10)(3/10)

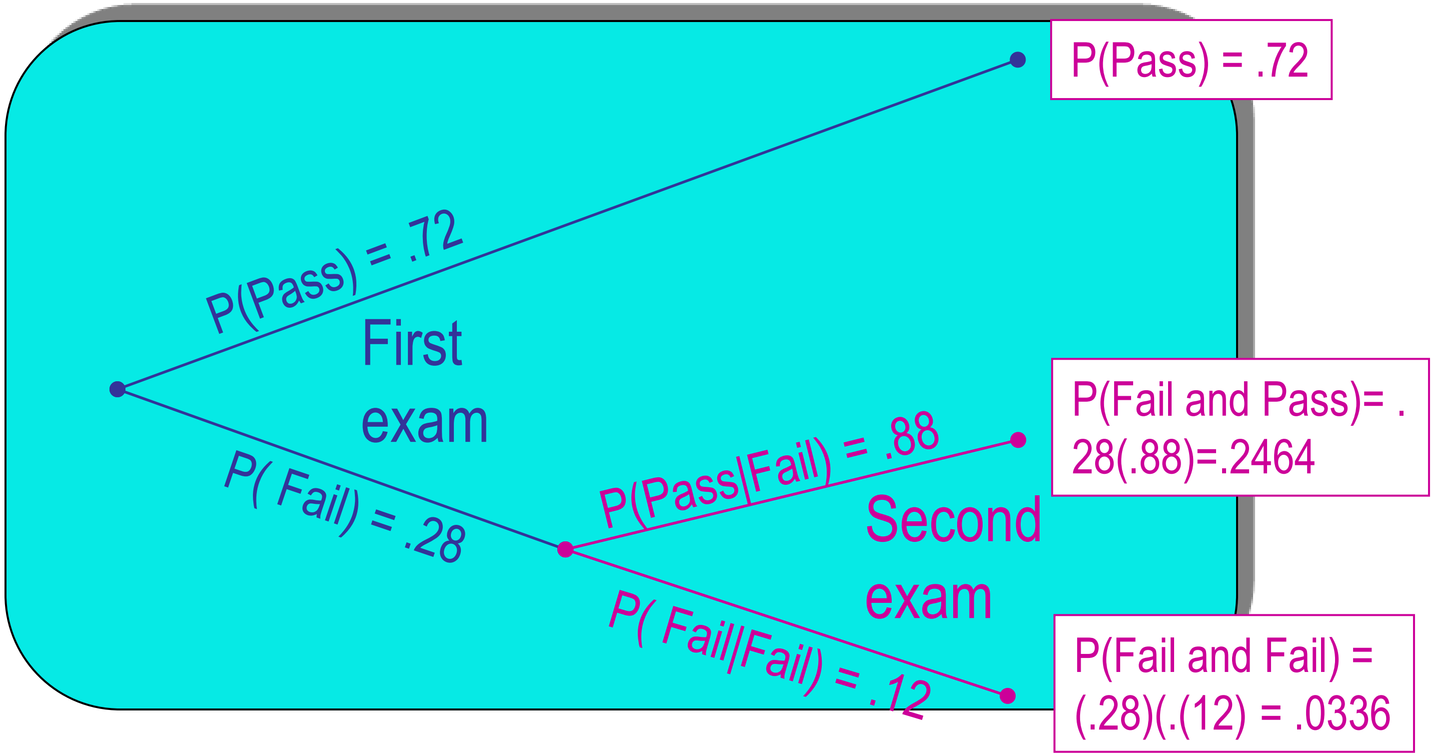
P(MM)=(7/10)(7/10)

Probability Trees

- Example 16 (conditional probabilities)
  - The pass rate of first-time takers for the bar exam at a certain jurisdiction is 72%.
  - Of those who fail, 88% pass their second attempt.
  - Find the probability that a randomly selected law school graduate becomes a lawyer (candidates cannot take the exam more than twice).

Probability Trees

- Solution



P(Pass) = P(Pass on first exam) + P(Fail on first and Pass on second) = .9664

Bayes' Law

- Conditional probability is used to find the probability of an event given that one of its possible causes has occurred.
- We use Bayes' law to find the probability of the possible cause given that an event has occurred.

Bayes' Formula

- Conditional probability is used to find the probability of an event given that one of its possible causes has occurred.
- We use Bayes' formula to find the probability of the possible cause given that an event has occurred.

$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{k} P(B \mid A_i)P(A_i)}, \quad j = 1, 2, \ldots, k$$

Bayes' Law

- Example 17
  - Medical tests can produce false-positive or false-negative results.
  - A particular test is found to perform as follows:
    - Correctly diagnose "Positive" 94% of the time.
    - Correctly diagnose "Negative" 98% of the time.
  - It is known that 4% of men in the general population suffer from the illness.
  - What is the probability that a man is suffering from the illness, if the test result were positive?

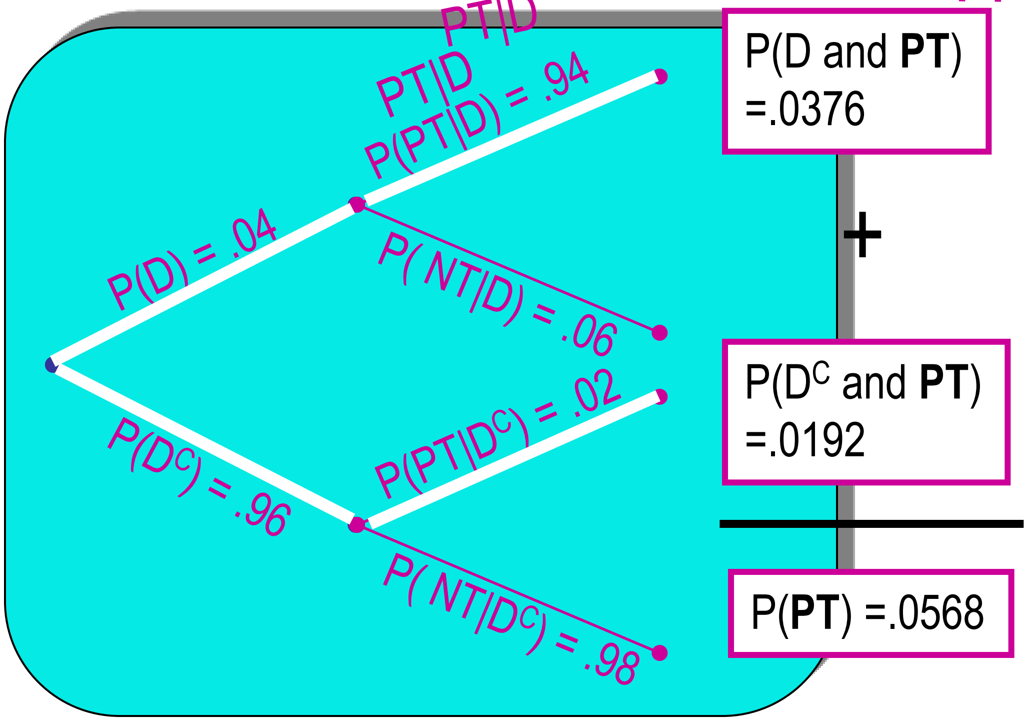
Bayes' Law

- Solution
  - Define the following events
    - D = Has a disease
    - D^C = Does not have the disease
    - PT = Positive test results
    - NT = Negative test results
  - Build a probability tree

Bayes' Law

- Solution – Continued
  - The probabilities provided are:
    - $P(D) = .04$ $P(D^C) = .96$
    - $P(PT|D) = .94$ $P(NT|D) = .06$
    - $P(PT|D^C) = .02$ $P(NT|D^C) = .98$
  - The probability to be determined is $\boxed{P(D|PT)}$

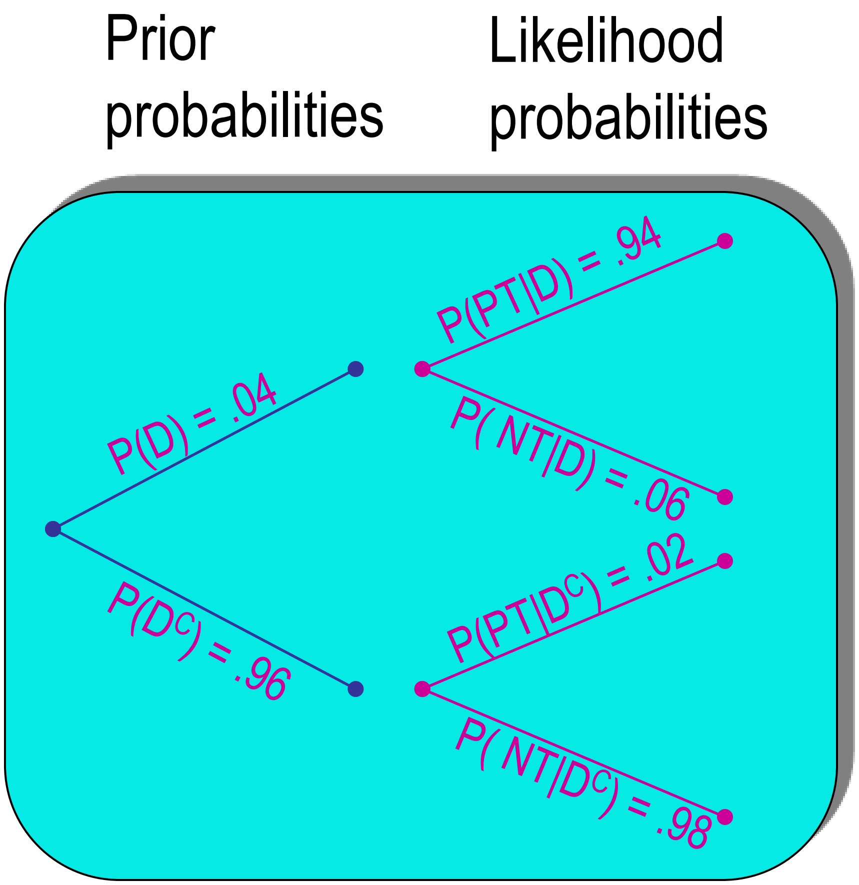
Bayes' Law



$$P(D\,|\,PT) = \frac{P(D \text{ and } PT)}{P(PT)}$$

$$= \frac{.0376}{.0568} = .6620$$

Bayes' Law



Posterior probabilities

$$P(D\,|\,PT) = \frac{.0376}{.0568} = .6620$$

EXAMPLE 18

An ice cream vendor sells three flavors: chocolate, strawberry, and vanilla. Forty five percent of the sales are chocolate, while 30% are strawberry, with the rest vanilla flavored. Sales are by the cone or the cup. The percentages of cones sales for chocolate, strawberry, and vanilla, are 75%, 60%, and 40%, respectively. For a randomly selected sale, define the following events:

A_1 = chocolate chosen

A_2 = strawberry chosen

A_3 = vanilla chosen

B = ice cream on a cone

B^C = ice cream in a cup

- Find the probability that the ice cream was sold on a cone and was

a) chocolate flavor

b) strawberry flavor

c) vanilla flavor

- **ANSWERS:**

a) $P(B \text{ and } A_1) = P(B/A_1).P(A_1) = (0.75)(0.45) = 0.3375$

b) $P(B \text{ and } A_2) = P(B/A_2).P(A_2) = (0.60)(0.30) = 0.18$

c) $P(B \text{ and } A_3) = P(B/A_3).P(A_3) = (0.40)(0.25) = 0.10$

- Find the probability that the ice cream was sold in a cup and was chocolate flavor

ANSWERS:

$P(B^C \text{ and } A_1) = P(B^C / A_1).P(A_1) = (0.25)(0.45) = 0.1125$

- Find the probability that the ice cream was sold on a cone.

ANSWER:

$P(B) = P(B \text{ and } A_1) + P(B \text{ and } A_2) + P(B \text{ and } A_3) = 0.3375 + 0.18 + 0.10 = 0.6175$

- Find the probability that the ice cream was sold in a cup.

ANSWER:

$P(B^C) = 1 – P(B) = 1 – 0.6175 = 0.3825$

- Find the probability that the ice cream was chocolate flavor, given that it was sold on a cone

ANSWER:

$P(A_1 / B) = P(A_1 \text{ and } B) / P(B) = 0.3375 / 0.6175 = 0.5466$

- Find the probability that the ice cream was chocolate flavor, given that it was sold in a cup

ANSWER:

$P(A_1 / B^C) = P(A_1 \text{ and } B^C) / P(B^C)$

$= 0.1125 / 0.3825 = 0.2941$

Random Variables and Discerete Probability Distributions

Random Variables and Probability Distributions

- A random variable is a function or rule that assigns a numerical value to each simple event in a sample space.
- A random variable reflects the aspect of a random experiment that is of interest for us.
- There are two types of random variables:
  - Discrete random variable
  - Continuous random variable.

Random Variables

- If X is a function that assigns a real numbered value to every possible event in a sample space of interest, X is called a random variable.
- It is denoted by capital letters such as X, Y and Z.
- The specified value of the random variable is unknown until the experimental outcome is observed.

EXAMPLES

- The experiment of flipping a fair coin. Outcome of the flip is a random variable.

$$S=\{H,T\}\rightarrow X(H)=1 \text{ and } X(T)=0$$

- Select a student at random from all registered students at METU. We want to know the weight of these students.

$$X = \text{the weight of the selected student}$$

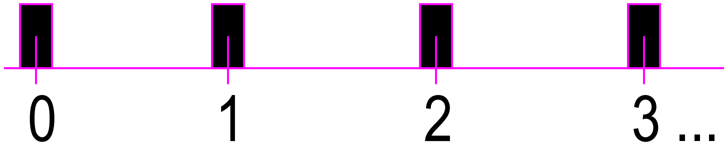
$$S: \{x: 45\text{kg} \leq X \leq 300\text{kg}\}$$

Discrete and Continuous Random Variables

- A random variable is discrete if it can assume a countable number of values.
- A random variable is continuous if it can assume an uncountable number of values.

Discrete random variable

After the first value is defined the second value, and any value thereafter are known.



Therefore, the number of values is countable

Continuous random variable

After the first value is defined, any number can be the next one

Therefore, the number of values is uncountable

Discrete Probability Distribution

- A table, formula, or graph that lists all possible values a discrete random variable can assume, together with associated probabilities, is called ***a discrete probability distribution***.

- To calculate the probability that the random variable **X** assumes the value x, P(**X** = x),
  - add the probabilities of all the simple events for which **X** is equal to x, or
  - Use probability calculation tools (tree diagram),
  - Apply probability definitions

Requirements for a Discrete Distribution

- If a random variable can assume values x_i, then the following must be true:

$$1.\ 0 \le p(x_i) \le 1 \text{ for all } x_i$$

$$2.\ \sum_{\text{all } x_i} p(x_i) = 1$$

EXAMPLE

- Consider an experiment in which a fair coin is tossed 3 times.

X = The number of heads

Let's assign 1 for head and 0 for tail. The sample space is

S: {TTT,TTH,THT,HTT,THH,HTH,HHT,HHH}

Possible values of X is 0, 1, 2, 3. Then, the probability distribution of X is

x	0	1	2	3	Total
p(x)	1/8	3/8	3/8	1/8	1

Distribution and Relative Frequencies

- In practice, often probability distributions are estimated from relative frequencies.
- **Example 7.1**
  - A survey reveals the following frequencies (1,000s) for the number of color TVs per household.

Number of TVs	Number of Households	x	p(x)
0	1,218	0	1218/Total = .012
1	32,379	1	.319
2	37,961	2	.374
3	19,387	3	.191
4	7,714	4	.076
5	2,842	5	.028
Total	**101,501**		**1.000**

Determining Probability of Events

- The probability distribution can be used to calculate the probability of different events
- **Example 7.1 – continued**

  Calculate the probability of the following events:

  – P(The number of color TVs is 3) = $P(X=3)$ =.191

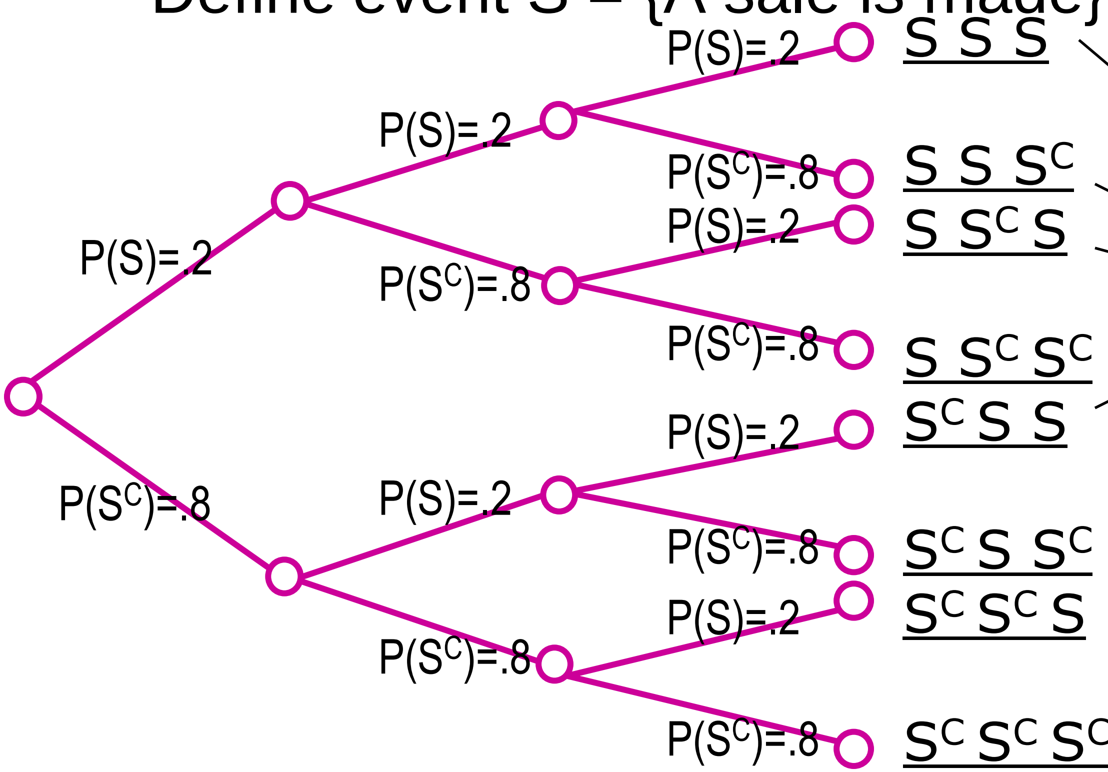
  – P(The number of color TVs is two or more) = $P(X\geq 2)=P(X=2)+P(X=3)+P(X=4)+P(X=5)=$ .374 +.191 +.076 +.028 = .669

Developing a Probability Distribution

- Probability calculation techniques can be used to develop probability distributions
- **Example 7.2**
  - A mutual fund sales person knows that there is 20% chance of closing a sale on each call she makes.
  - What is the probability distribution of the number of sales if she plans to call three customers?

Developing a Probability Distribution

- Solution
  - Use probability rules and trees
  - Define event S = {A sale is made}.



(.2)(.2)(.8)= .032

X	**P(x)**
3	$.2^3 = .008$
2	$3(.032)=.096$
1	$3(.128)=.384$
0	$.8^3 = .512$

The Cumulative Distribution Function

- If X is a random variable, then the cumulative distribution function (cdf), denoted by F(x) is given by

$$F(x) = P(X \le x) = \sum_{y:y \le x} p(y)$$

for all real numbers x. It is a non-decreasing step function of x and it is right continuous.

The Cumulative Distribution Function

- For any two numbers a and b, $a \leq b$

$$P(a \leq X \leq b) = F(b) - F(a^-)$$

where a^- represents the largest possible X value which is less than a.

- If a and b are integers,

$$P(a \leq X \leq b) = F(b) - F(a-1)$$

- Taking a=b,

$$P(X=a) = F(a) - F(a-1)$$

EXAMPLE

- Let X is the number of days of sick leave taken by a randomly selected employee of a large company during a particular year. If the max. number of allowable sick days per year is 14, possible values of X are 0,1,2,…,14. With F(0)=0.58, F(1)=0.72, F(2)=0.76,F(3)=0.81,F(4)=0.88 and F(5)=0.94, find

$P(2 \leq X \leq 5) =$

$P(X = 3) =$

$P(X \leq 2) =$