IE2151
Probability for Industrial Engineers
Problem Solving Sessions

Session 06 – Probability

Definitions in Probability

- Outcome is a possible observation in an experiment
- Experiment is any action or process to collect outcomes
- **Random experiment** is any action or process that gives different outcomes under the same conditions.
  - Tossing a coin, drawing cards from a deck, rolling a die, etc.
- **Sample Space** is the set of all possible outcomes of the experiment.
  - *S* is discrete if it consists of a finite or countable infinite set of outcomes.
  - *S* is continuous if it contains an interval of real numbers.

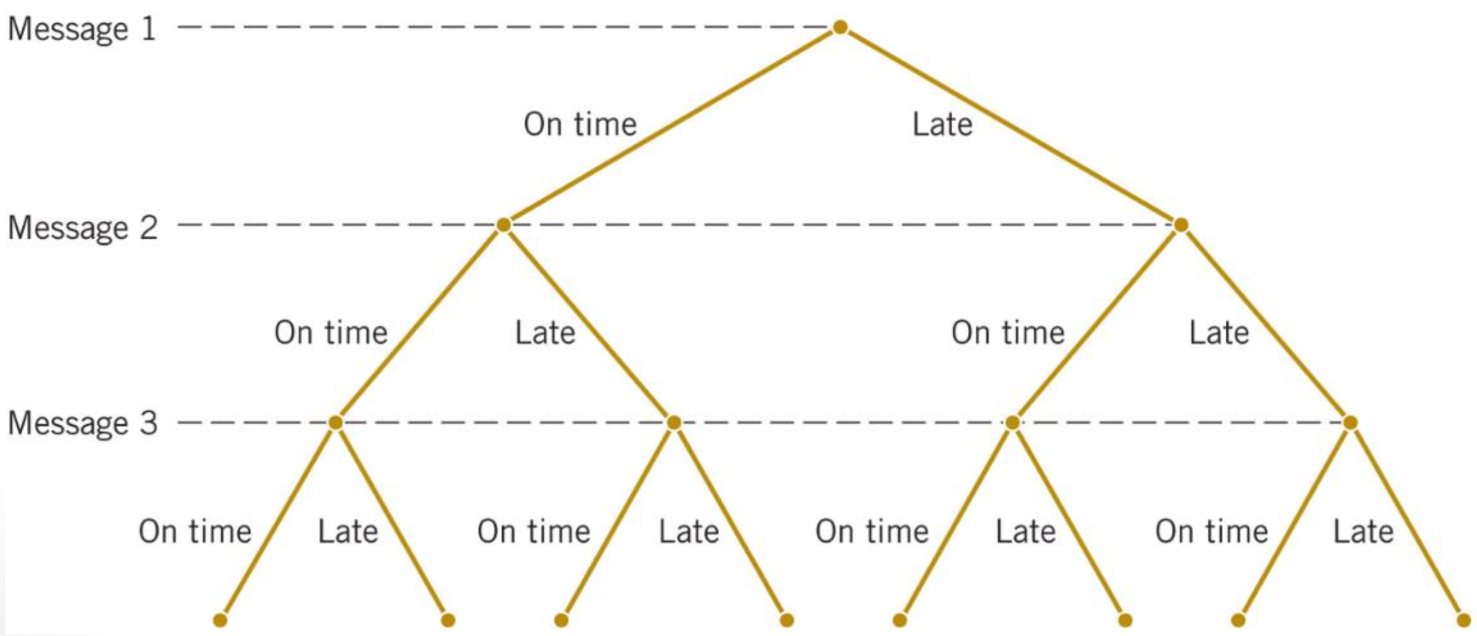
Ex. 2-1: Defining Sample Spaces

- Randomly select a camera and record the recycle time of a flash.
  - $S = R^+ = \{x \mid x > 0\}$, the positive real numbers.
- Suppose it is known that all recycle times are between 1.5 and 5 seconds. Then
  - $S = \{x \mid 1.5 < x < 5\}$ is continuous.
- It is known that the recycle time has only three values(low, medium or high). Then
  - $S = \{low, medium, high\}$ is discrete.
- Does the camera conform to minimum recycle time specifications?
  - $S = \{yes, no\}$ is discrete.

Ex. 2-2: Sample Space Defined By A Tree Diagram

- Messages are classified as on-time(o) or late(l). Classify the next 3 messages.

$$S = \{ooo, ool, olo, oll, loo, lol, llo, lll\}$$



Definitions in Probability

Example:

1. Flip a coin 3 times, Observe the sequence of heads/tails

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

2. Flip a coin 3 times, Observe # of heads

   {0, 1, 2, 3}

Event

- Set of outcomes (Must know all outcomes)
- Event $\subset$ Sample Space
  - Simple (or elementary) event*:* consists of exactly one outcome
  - Compound event: consists of more than one outcome

Event Examples

- **For an experiment:** Roll a die, observe the shown numbers

- **Outcomes:**

  number = 1,2,3,4,5,6

- **Sample space:**

  S = {1,2,3,4,5,6}

- **Event examples:**
  - Simple event: E = {4}
  - Compound event :
    - E_1 = {number < 3} = {1,2}
    - E_2 = {number is odd} = {1,3,5 }

Event Combinations

- The **Union** of two events consists of all outcomes that are contained in one event <u>or</u> the other, denoted as $E_1 \cup E_2$.

- The **Intersection** of two events consists of all outcomes that are contained in one event <u>and</u> the other, denoted as $E_1 \cap E_2$.

- The **Complement** of an event is the set of outcomes in the sample space that are <u>not</u> contained in the event, denoted as E' or $\bar{E}$.

Ex. 2-3: Discrete Events

- Suppose that the recycle times of two cameras are recorded. Consider only whether or not the cameras conform to the manufacturing specifications. We abbreviate *yes* and *no* as *y* and *n*. The sample space is $S = \{yy, yn, ny, nn\}$.
- Suppose, E_1 denotes an event that at least one camera conforms to specifications, then $E_1 = \{yy, yn, ny\}$
- Suppose, E_2 denotes an event that no camera conforms to specifications, then $E_2 = \{nn\}$
- Suppose, E_3 denotes an event that at least one camera <u>does not</u> conform, then $E_3 = \{yn, ny, nn\}$.
  - Then $E_1 \cup E_3 = S$
  - Then $E_1 \cap E_3 = \{yn, ny\}$
  - Then $E_1' = \{nn\}$

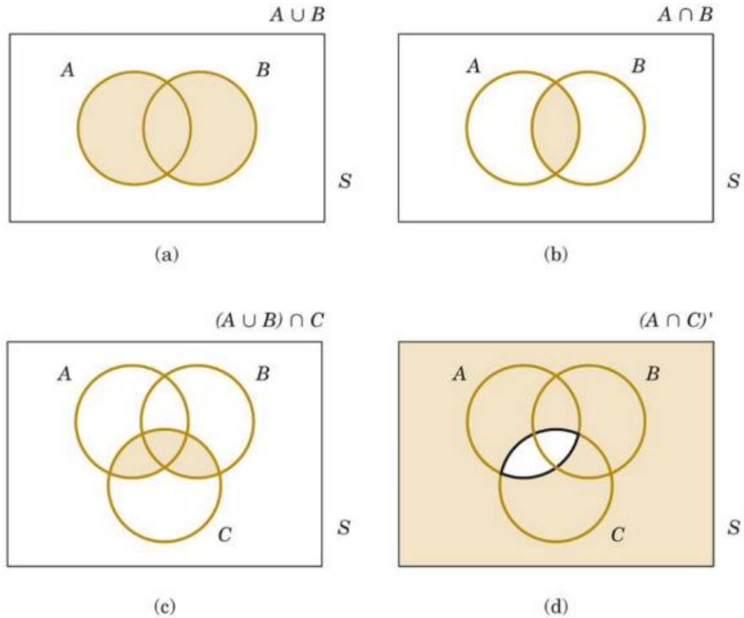
Ex. 2-4: Continous Events

- Measurements of the thickness of a part are modeled with the sample space: $S = R^+$.
- Let $E_1 = \{x \mid 10 \le x < 12\}$,
- Let $E_2 = \{x \mid 11 < x < 15\}$
  - Then $E_1 \cup E_2 = \{x \mid 10 \le x < 15\}$
  - Then $E_1 \cap E_2 = \{x \mid 11 < x < 12\}$
  - Then $E_1' = \{x \mid 0 < x < 10 \text{ or } x \ge 12\}$
  - Then $E_1' \cap E_2 = \{x \mid 12 \le x < 15\}$

Set vs. Probability

Set Algebra	Probability
Set	Event
Universal set	Sample space
Element	Outcome

- Events *A* & *B* contain their respective outcomes. The shaded regions indicate the event relation of each Venn diagram.



Probability Concept

- Probability allows us to quantify the likelihood associated with uncertain events, that is, events that result from random experiments.
- Probabilities are reported as
  - proportions (between 0 and 1)
  - percentages (between 0% and 100%).
- Thus the statements
  - P(A)=.30, the probability of event A occurring is .30, and
  - the event A has a 30% chance of occurring
- are equivalent.

Probability of an Event

- **P( )** is a function that maps the event in the sample space to a real number.

- From the experiment: Roll a die
- **Outcomes:**

  number = 1,2,3,4,5,6

- **Sample space:**

  $S = \{1,2,3,\ldots,6\}$

- **Event examples**:

  $E_1 = \{\text{number} < 3\} = \{1,2\}$
  $E_2 = \{\text{number is odd}\} = \{1,3,5\}$

Probability of an Event

- There are several ways to determine probability of an event:
- As relative frequencies of occurrence;
  - Repeat the experiment
  - Calculate the relative frequency of the occurrence of the event of interest
- By assuming that events are equal likely
  - Die, coin, etc.
- From subjective estimates.
  - To find the probability that a horse will win a race,
    - Previous records of all the horses entered in the race
    - The records of the jockeys riding the horse

Assign Probability Value to an Event

- **Example**
- Purpose: To find the probability that a product can be defective
- Experiment: Each product in a production line is checked to determine whether it is defective or not.
- There are two consequences into an experiment
  - DF: Defective
  - ND: Non defective
- Let
  - n : The number of experiments which were conducted
  - n(DF) : The number of defective products into *n* experiments
- Using first approach, i.e. relative frequency of occurrence

$$P(DF) = n(DF)/n$$

- As n grows large, n(DF)/n ratio converges to a steady number, called the limiting relative frequency, which is used to estimate P(DF).

Assign Probability Value to an Event

- **Example**
- Check the 600 products (n)
- Classify each of them into two classes as defective and non-defective
- Find the number of defective products
  - n(DF) = 60
- Estimate the probability that a product is a defective

$$
\begin{aligned}
P(DF) &= n(DF) / n \\
&= 60 / 600 \\
&= 0.10
\end{aligned}
$$

- The product is defective with the probability of 0.10

Assign Probability Value to an Event

- If the sample space for an experiment contains *N* elements, all of which are equally likely occur,
- Such as
  - rolling a die, $S = \{1,2,3,4,5,6\}$
  - tossing a coin, $S = \{H,T\}$
  - the probability of each of the *N* points is equal and $1/N$.
- If an experiment can result in any one of *N* different equally likely outcomes, and if exactly *n* of these outcomes correspond to event A, then the probability of event A is

$$P(A) = \frac{n}{N}$$

Example 1 (Walpole, p. 41)

- A statistics class for engineers consists of 25 industrial, 10 mechanical, 10 electrical, and 8 civil engineering students.
- If a person is randomly selected by the instructor to answer a question, find the probability that the student chosen is
  a) An industrial engineering major
  b) A civil engineering or an electrical engineering major

- **Solution:**

Events: I, M, E, C

a) $P(I) = \dfrac{n(I)}{n} = \dfrac{25}{53}$

b) $P(C \cup E) = P(C) + P(E) - P(C \cap E)$

$= \dfrac{8+10}{53} = \dfrac{18}{53}$

Example 1

- **RStudio output:**

```
> #Define number of students in each department.
> students<-c(I=25,M=10,E=10,C=8)
>
> #Calculate the probabilities of each event.
> probs<-prop.table(students); probs
        I         M         E         C
0.4716981 0.1886792 0.1886792 0.1509434
>
> #Find the probability that the chosen student
> #is an industrial eng. major.
> probs["I"]
        I
0.4716981
>
> #Find the probability that the chosen student
> #is a civil eng. or an electrical eng. major.
> sum(probs[c("C","E")])
[1] 0.3396226
```

Ex. 2-10: Semiconductor Wafers

- A wafer is randomly selected from a batch that is classified by contamination and location.

Contamination	Location of Tool		Total
	Center	Edge	
Low	514	68	582
High	112	246	358
Total	626	314	940

- Let H be the event of high concentrations of contaminants. Then $P(H) = 358/940$.
- Let C be the event of the wafer being located at the center of a sputtering tool. Then $P(C) = 626/940$.
- $P(H \cap C) = 112/940$
- $P(H \cup C) = P(H) + P(C) - P(H \cap C) = (358 + 626 - 112)/940$

  This is the addition rule.

Ex. 2-10: Semiconductor Wafers

- **RStudio output:**

```
> #Create contamination vectors of low and high.
> low<-c(center=514,edge=68)
> high<-c(center=112,edge=246)
>
> #Create a data frame by combining them as rows.
> contamination<-rbind(low,high)
>
> #Create a probability table.
> probs<-prop.table(contamination); probs
        center      edge
low  0.5468085 0.07234043
high 0.1191489 0.26170213
>
> #Calculate the probabilies.
> p_H<-sum(probs["high",]); p_H
[1] 0.3808511
> p_C<-sum(probs[,"center"]); p_C
[1] 0.6659574
> p_HnC<-probs["high","center"]; p_HnC
[1] 0.1191489
> p_HuC<-p_H+p_C-p_HnC; p_HuC
[1] 0.9276596
```

Example 2 (Walpole, p. 40)

- A die is loaded in such a way that an even number is twice as likely to occur as an odd number.

  a) If E is the event that a number less than 4 occurs on a single toss of the die, P(E)=?

  b) Let A be the event that an even number turns up and let B be the event that a number divisible by 3 occurs. $P(A \cup B) = ?$ and $P(A \cap B) = ?$

Example 2 (Walpole, p. 40)

- **Solution:**

$S = \{1,2,3,4,5,6\}$

Let k = prob. that odd number occurs

Let $2k$ = " " even " "

$9k = 1 \Rightarrow k = 1/9$

a) $E = \{1,2,3\}$ $\quad P(E) = 2 \cdot \frac{1}{9} + 1 \cdot \frac{2}{9} = \frac{4}{9}$

b) $A = \{2,4,6\}$ $\quad B = \{3,6\}$

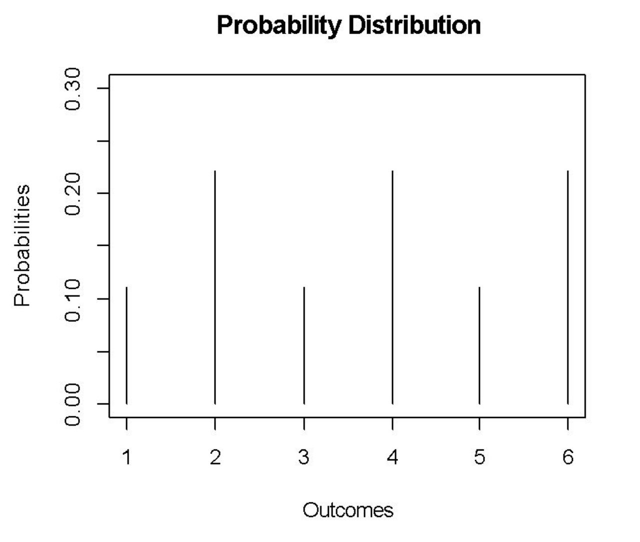
$A \cup B = \{2,3,4,6\}$ $\quad A \cap B = \{6\}$

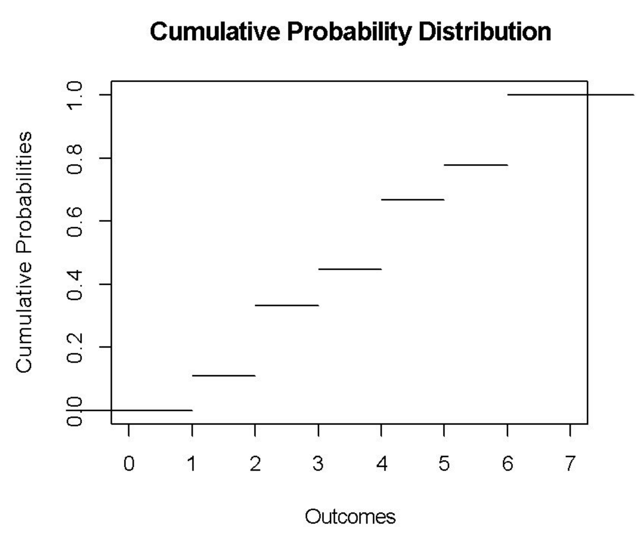
$P(A \cup B) = 1 \cdot \frac{1}{9} + 3 \cdot \frac{2}{9} = \frac{7}{9}$ $\quad P(A \cap B) = \frac{2}{9}$

Example 2

- **RStudio output:**

```
> #Generate the vector of probabilities
> probs<-rep(c(1/9,2/9),3)
> #Plot the probabilities
> plot(probs, type='h',ylim=c(0,0.3),
+      xlab = "Outcomes", ylab="Probabilities",
+      main = "Probability Distribution")
>
> #Generate the vector of cumulative probabilities
> cumprobs <- cumsum(probs)
>
> #Plot the cumulative probabilites
> steps<-stepfun(1:6,c(0,cumprobs))
> plot(steps, verticals=F, col.points = 'transparent',
+      xlab = "Outcomes",
+      ylab="Cumulative Probabilities",
+      main = "Cumulative Probability Distribution")
>
> #Calculate the probabilities.
> p_E<-sum(probs[1:3]); p_E
[1] 0.4444444
> p_AuB<-sum(probs[c(2,3,4,6)]); p_AuB
[1] 0.7777778
> p_AnB<-sum(probs[6]); p_AnB
[1] 0.2222222
```





Example 3 (Walpole, p. 42)

- In a poker hand consists of 5 cards, find the probability of holding 2 aces and 3 jacks?

- **Solution:**

$$\binom{52}{5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2} = 2{,}598{,}960$$

* 52 cards in a deck, choose 5

* 4 aces, choose 2: $\binom{4}{2} = \frac{4 \cdot 3}{2} = 6$ ways

* 4 jacks, choose 3: $\binom{4}{3} = 4$ ways

$6 \cdot 4 = 24$ ways

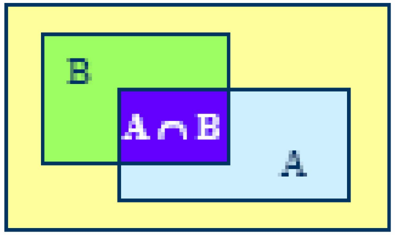
$$P\left(\begin{matrix}\text{2 aces and} \\ \text{3 jacks}\end{matrix}\right) = \frac{24}{2{,}598{,}960} = \frac{1}{108{,}290} = 0.00000923$$

```
> choose(4,2)*choose(4,3)/choose(52,5)
[1] 9.234463e-06
```

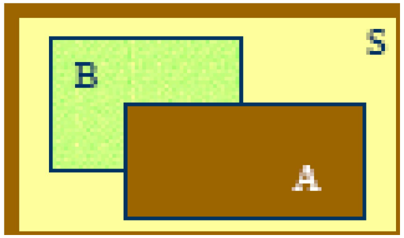
Conditional Probability

- Let A and B two events with $P(B) > 0$.
- **The conditional probability of A occurring given that event B has already occurred** is denoted by $P(A|B)$ and can be calculated from the formula

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$



$$P(A \mid S) = \frac{P(A \cap S)}{P(S)} = P(A)$$



Example 4

- Let’s consider a group of 100 people of whom;
  - 40 are college graduates
  - 20 are self employed
  - 10 are both college graduates & self-employed
- Let
  - B represent the set of college graduates
  - A represent the set of self-employed
  - $A \cap B$ is the set of college graduates who are self employed
- From the group of 100 , one person is about to be randomly selected.
- Calculate the following probabilities:
  - The selected person is a college graduate
  - The selected person is self-employed
  - The selected person is both a college graduate and self employed
  - The selected person is a college graduate given that he/she is self-employed
  - The selected person is self-employed given that he/she is a college graduate

Example 4

- **Solution:**

$$P(A) = \frac{20}{100} = 0.2$$

$$P(B) = \frac{40}{100} = 0.4$$

$$P(A \cap B) = \frac{10}{100} = 0.1$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.4} = 0.25$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.1}{0.2} = 0.5$$

Example 4

- **RStudio otuput:**

```
> #Define a function called condprob.
> condprob <- function(pab,pb){
+   return(pab/pb)
+ }
>
> #Calculate the probabilities
> p_b<-40/100; p_b
[1] 0.4
> p_a<-20/100; p_a
[1] 0.2
> p_anb<-10/100; p_anb
[1] 0.1
> p_aub<-p_a+p_b-p_anb; p_aub
[1] 0.5
> condprob(p_anb,p_a)
[1] 0.5
> condprob(p_anb,p_b)
[1] 0.25
```

Conditional Probability

- $$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) > 0$$

- The conditional probability satisfies the required properties of probability theory. That is,

1. $0 \leq P(A|B) \leq 1$
2. $P(S|B) = 1$
3. $P(A_1 \cup A_2 \cup \cdots \cup A_k|B) = P(A_1|B) + P(A_2|B) + \cdots + P(A_k|B)$

Example 5 (Schaum's outline series, p. 62)

- In a certain college, 25% of the students failed math, 15% of the students failed chemistry, 10% of the students failed both math and chemistry. A student is selected at random.

a) If he failed chemistry what is the probability that he failed math?

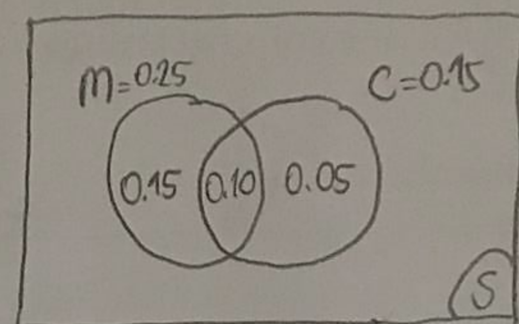
b) If he failed math, what is the probability that he failed chemistry?

c) What is the probability that he failed math or chemistry?

Example 5 (Schaum's outline series, p. 62)

- **<u>Solution:</u>**

Events: M = the student failed math
C = the student failed chemistry



$$\left.\begin{array}{l} P(M)=0.25 \\ P(C)=0.15 \end{array}\right\} P(M\cap C)=0.10$$

a) $P(M|C)=\dfrac{P(M\cap C)}{P(C)}=\dfrac{0.10}{0.15}=\dfrac{2}{3}$

b) $P(C|M)=\dfrac{P(C\cap M)}{P(M)}=\dfrac{0.10}{0.25}=\dfrac{2}{5}$

c) $P(M\cup C)=P(M)+P(C)-P(M\cap C)$
$=0.25+0.15-0.10=0.30$

Example 5

- **RStudio otuput:**

```
> p_m<-0.25
> p_c<-0.15
> p_mnc<-0.10
>
> #Part a)
> condprob(p_mnc,p_c)
[1] 0.6666667
>
> #Part b)
> condprob(p_mnc,p_m)
[1] 0.4
>
> #Part c)
> p_m+p_c-p_mnc
[1] 0.3
```

Example 6

- The probability that a regularly scheduled flight departs on time is $P(D)=0.83$; the probability that it arrives on time is $P(A)=0.82$; and the probability that it departs and arrives on time is $P(D \cap A)= 0.78$.
- Find the probability that a plane
  a) Arrives on time given that it departed on time
  b) Departed on time given that it has arrived on time.
  c) Arrives on time given that it did not depart on time.

P(it did not depart on time) = $1 – P(D) = 1- 0.83=0.17$

Example 6

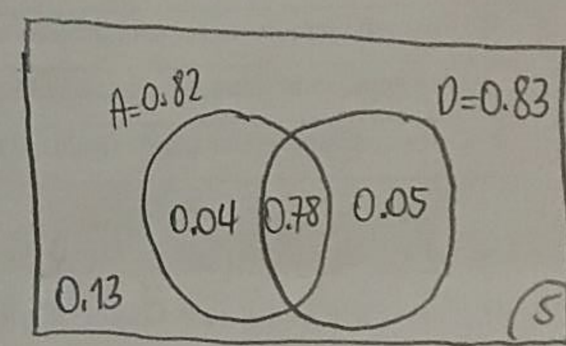
- **Solution:**

```
> p_d<-0.83
> p_a<-0.82
> p_dna<-0.78
>
> #Part a)
> condprob(p_dna,p_d)
[1] 0.939759
> #Part b)
> condprob(p_dna,p_a)
[1] 0.9512195
> #Part c)
> condprob(p_a-p_dna,1-p_d)
[1] 0.2352941
```

a) $P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{0.78}{0.83} \cong 0.9398$

b) $P(D|A) = \frac{P(A \cap D)}{P(A)} = \frac{0.78}{0.82} \cong 0.9512$

c) $P(A|D') = \frac{P(A \cap D')}{P(D')} = \frac{0.04}{0.17} \cong 0.2353$



Example 7

- There are three columns entitled “Art” (A), “Books” (B) and “Cinema” (C) in a new magazine. Reading habits of a randomly selected reader with respect to chosen columns are:

read regularly	A	B	C	$A \cap B$	$A \cap C$	$B \cap C$	$A \cap B \cap C$
probability	.14	.23	.37	.08	.09	.13	.05

a) What is the probability of column A given that they read column B?

b) What is the probability of reading A given that they are reading B or C columns?

c) What is the probability of reading column A given that they are reading at least one column?

d) What is the probability of reading A or B columns given that they read C columns?

Example 7

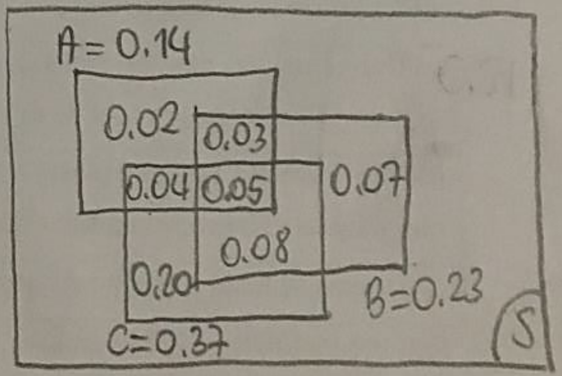
- **Solution:**

a) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.08}{0.23} = 0.348$

b) $P(A|B \cup C) = \dfrac{P(A \cap (B \cup C))}{P(B \cup C)} = \dfrac{0.12}{0.47} = 0.255$

c) $P(A|A \cup B \cup C) = \dfrac{P(A)}{P(A \cup B \cup C)} = \dfrac{0.14}{0.49} = 0.286$

d) $P(A \cup B|C) = \dfrac{P((A \cup B) \cap C)}{P(C)} = \dfrac{0.17}{0.37} = 0.459$



Example 7

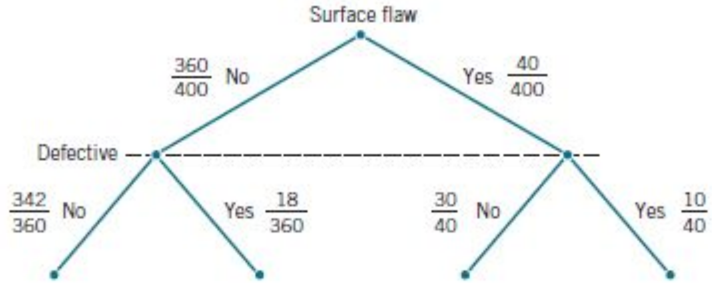
- **RStudio otuput:**

```
> p_a<-0.14
> p_b<-0.23
> p_c<-0.37
> p_anb<-0.08
> p_anc<-0.09
> p_bnc<-0.13
> p_anbnc<-0.05
>
> #Part a)
> condprob(p_anb,p_b)
[1] 0.3478261
>
> #Part b)
> p_buc<-p_b+p_c-p_bnc; p_buc
[1] 0.47
> #We have p_an(buc)=p_anb+p_anc-p_anbnc
> p_anbuc<-p_anb+p_anc-p_anbnc; p_anbuc
[1] 0.12
> condprob(p_anbuc,p_buc)
[1] 0.2553191
>
> #Part c)
> p_aubuc<-p_a+p_b+p_c-p_anb-p_anc-p_bnc+p_anbnc; p_aubuc
[1] 0.49
> condprob(p_a,p_aubuc)
[1] 0.2857143
>
> #Part d)
> #We have p_(aub)nc=p_anc+p_bnc-p_anbnc
> p_aubnc<-p_anc+p_bnc-p_anbnc; p_aubnc
[1] 0.17
> condprob(p_aubnc,p_c)
[1] 0.4594595
```

Ex. 2-11: Surface Flaws and Defectives

- Table provides an example of 400 parts classified by surface flaws and as (functionally) defective.
- Let D denote the event that a part is defective, and let F denote the event that a part has a surface flaw.
- There are 4 probabilities conditioned on flaws as seen below:

Parts Classified			
Defective	Surface Flaws		Total
	Yes (F)	No (F')	
Yes (D)	10	18	28
No (D')	30	342	372
Total	40	360	400



$$P(F) = 40/400 \text{ and } P(D) = 28/400$$

$$P(D \mid F) = P(D \cap F)/P(F) = \tfrac{10}{400} \Big/ \tfrac{40}{400} = \tfrac{10}{40}$$

$$P(D' \mid F) = P(D' \cap F)/P(F) = \tfrac{30}{400} \Big/ \tfrac{40}{400} = \tfrac{30}{40}$$

$$P(D \mid F') = P(D \cap F')/P(F') = \tfrac{18}{400} \Big/ \tfrac{360}{400} = \tfrac{18}{360}$$

$$P(D' \mid F') = P(D' \cap F')/P(F') = \tfrac{342}{400} \Big/ \tfrac{360}{400} = \tfrac{342}{360}$$

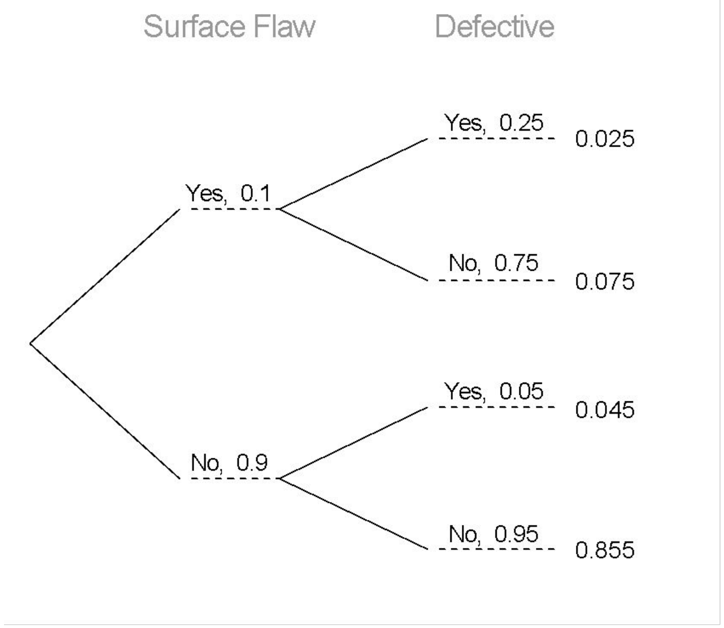
Ex. 2-11: Surface Flaws and Defectives

- **RStudio otuput:**

```
> #Create contamination vectors of low and high.
> d<-c(f=10,fprime=18)
> dprime<-c(f=30,fprime=342)
>
> #Create a data frame by combining them as rows.
> flaws<-rbind(d,dprime)
>
> #Create a probability table.
> probs<-prop.table(flaws); probs
          f fprime
d      0.025  0.045
dprime 0.075  0.855
>
> #Calculate the probabilies.
> p_d<-sum(probs[1,]); p_d
[1] 0.07
> p_dprime<-sum(probs[2,]); p_dprime
[1] 0.93
> p_f<-sum(probs[,1]); p_f
[1] 0.1
> p_fprime<-sum(probs[,2]); p_fprime
[1] 0.9
```

Ex. 2-11: Surface Flaws and Defectives

- **RStudio otuput:**




```
> #Calculate the conditional probabilies.
> condprobs<-probs
> for(i in 1:2)
+   for(j in 1:2)
+   condprobs[i,j]<-condprob(probs[i,j],colSums(probs)[j])
> condprobs
         f fprime
d      0.25   0.05
dprime 0.75   0.95
>
> #Install and load the package openintro to create tree diagrams.
> install.packages("openintro")
> library(openintro)
> old<-options(digits=4)
> treeDiag(c("Surface Flaw","Defective"),
      colSums(probs),
      list(condprobs[,"f"],condprobs[,"fprime"]))
> options(old)
```

Ex. 2-12: Conditional Probability and Random Sampling

- A batch of 50 parts contains 10 made by Tool 1 and 40 made by Tool 2. If 2 parts are selected randomly*,

a) What is the probability that the 2nd part came from Tool 2, given that the 1st part came from Tool 1?

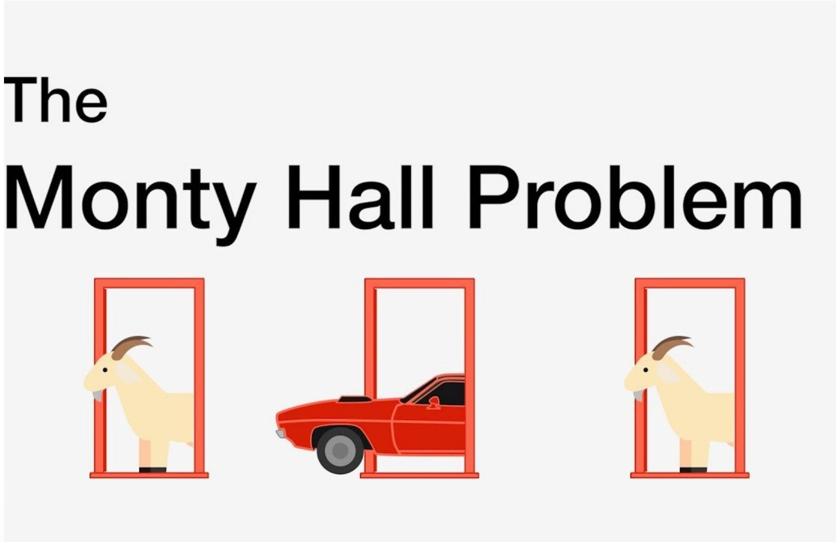
b) What is the probability that the 1st part came from Tool 1 and the 2nd part came from Tool 2?

- **Solution:**
- Part a)
  - $P(E_1) = P(1^{st}$ part came from Tool 1$) = 10/50$
  - $P(E_2 \mid E_1) = P(2^{nd}$ part came from Tool 2 given that 1^{st} part came from Tool 1$)$
    $= 40/49$

- Part b)
  - $P(E_1 \cap E_2) = P(1^{st}$ part came from Tool 1 and 2^{nd} part came from Tool 2$)$
    $= (10/50)\cdot(40/49) = 8/49$

*Selected randomly implies that at each step of the sample, the items remain in the batch are equally likely to be selected.

Question??



- 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 No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

References

- Altıparmak, F., (2005), Lecture Notes on Probability, http://w3.gazi.edu.tr/~fulyaal
- Walpole, Myers, Myers, Ye, (2002), Probability & Statistics for Engineers & Scientists.
- Montgomery, D. C., & Runger, G. C. (2006). *Applied statistics and probability for engineers.* 4th Edition. Hoboken, NJ: Wiley.