

Statistical and Mathematical Methods for Data Analysis

Dr. Faisal Bukhari

**Punjab University College of Information Technology
(PUCIT)**

Textbooks

❑ **Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer

❑ **Elementary Statistics: Picturing the World**, 6th Edition, Ron Larson and Betsy Farber

❑ **Elementary Statistics**, 13th Edition, Mario F. Triola

Reference books

- ❑ **Probability and Statistical Inference, Ninth Edition,** Robert V. Hogg, Elliot A. Tanis, Dale L. Zimmerman
- ❑ **Probability Demystified,** Allan G. Bluman
- ❑ **Practical Statistics for Data Scientists: 50 Essential Concepts,** Peter Bruce and Andrew Bruce
- ❑ **Schaum's Outline of Probability,** Second Edition, Seymour Lipschutz, Marc Lipson
- ❑ **Python for Probability, Statistics, and Machine Learning,** José Unpingco

References

Readings for these lecture notes:

❑ **Schaum's Outline of Probability, Second Edition**
(Schaum's Outlines)

by by Seymour Lipschutz, Marc Lipson

❑ **Probability & Statistics for Engineers & Scientists**,
Ninth Edition, Ronald E. Walpole, Raymond H.
Myer

❑ **Introduction to Probability** SECOND EDITION
Dimitri P. Bertsekas and John N. Tsitsiklis

These notes contain material from the above resources.

Recall: Mutually Exclusive or Disjoint

Two events A and B are **mutually exclusive, or disjoint**, if $A \cap B = \{ \}$ or \emptyset

OR

Two events A and B are **mutually exclusive, or disjoint**, if $A \cap B = \emptyset$, that is, if A and B have no elements in common.

OR

Events **A** and **B** are **disjoint** (or **mutually exclusive**) if they cannot occur at the same time. (That is, disjoint events do not overlap.)

Addition Rule I

Addition Rule I: When two events are **mutually exclusive**,

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

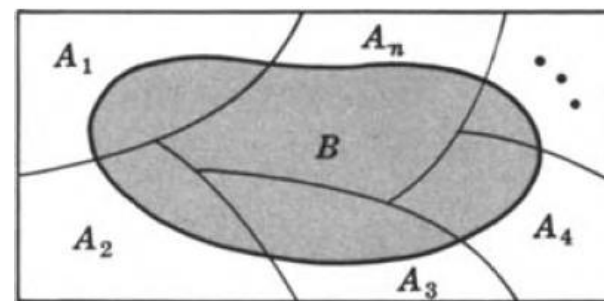
OR

$$P(A \cup B) = P(A) + P(B)$$

PARTITIONS AND BAYES' THEOREM

Suppose the events A_1, A_2, \dots, A_n form a partition of a sample space S ; that is, the events A_i are **mutually exclusive** and their union is S . Now let B be any other event. Then

$$\begin{aligned} B &= S \cap B = (A_1 \cup A_2 \cup \dots \cup A_n) \cap B \\ &= (A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B) \end{aligned}$$



where the $A_i \cap B$ are also **mutually exclusive**.

$$\therefore P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B)$$

Using the **multiplication theorem**, we get

$$\therefore P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n) \dots \dots \dots (1)$$

Total probability

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

The above formula is sometimes called the **theorem of total probability** or the **rule of elimination**.

On the other hand, for any i , the **conditional probability** of A_i given B is defined by

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} \dots\dots\dots(2)$$

Substitute value of **$P(B)$** from (1) and

$P(A_i \cap B) = P(A_i)P(B|A_i)$ in (2), we get

$$\therefore P(A_i|B) =$$

$$\frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)}$$

Bayes' Theorem

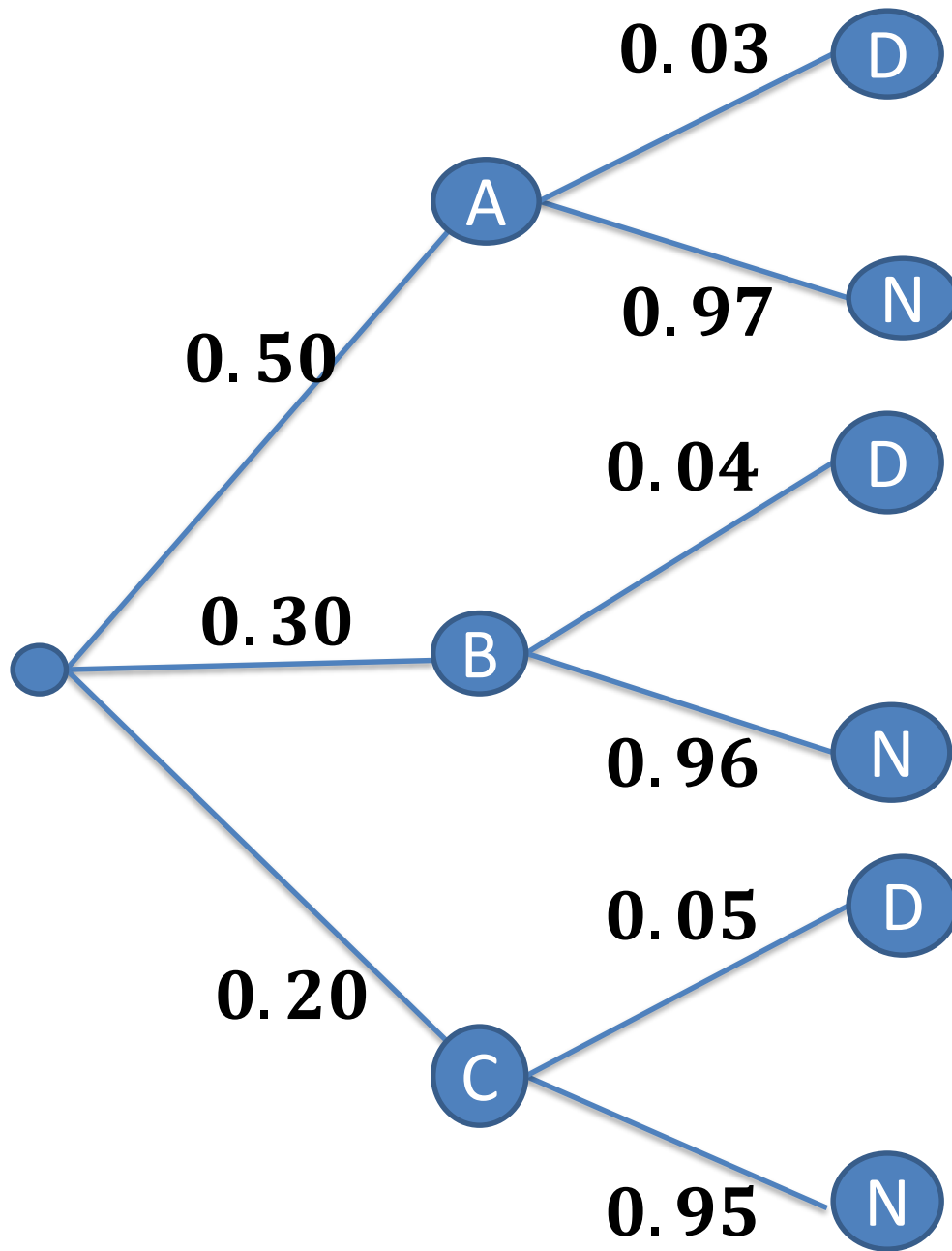
Suppose A_1, A_2, \dots, A_n is a partition of S and B is any event. Then for any i ,

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{P(A_1)P(B | A_1) + P(A_2)P(B | A_2) + \dots + P(A_n)P(B | A_n)}$$

or

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_{i=1}^n P(A_i)P(B | A_i)}$$

Example: Three machines **A**, **B** and **C** produce respectively **50%**, **30%** and **20%** of the total number of items of a factory. The percentages of **defective output** of these machines are **3%**, **4%** and **5%**. If an item is selected at random, find the probability that the item is **defective**.



Solution:

Let **D** be the event that the item is **defective**

$$\therefore P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)$$

$$\begin{aligned}\therefore \mathbf{P(D)} &= \mathbf{P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C)} \\ &= (0.50)(0.03) + (0.30)(0.04) + (0.20)(0.05) \\ &= \mathbf{0.037} \quad \mathbf{(or\ 3.7\%)}\end{aligned}$$

Example: Consider the factory in the preceding example. Suppose an item is selected at random and is found to be **defective**. Find the probability that the item was produced by **machine A**; that is, find **$P(A|D)$** .

Solution:

By Bayes' theorem,

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_{i=1}^n P(A_i)P(B | A_i)}$$

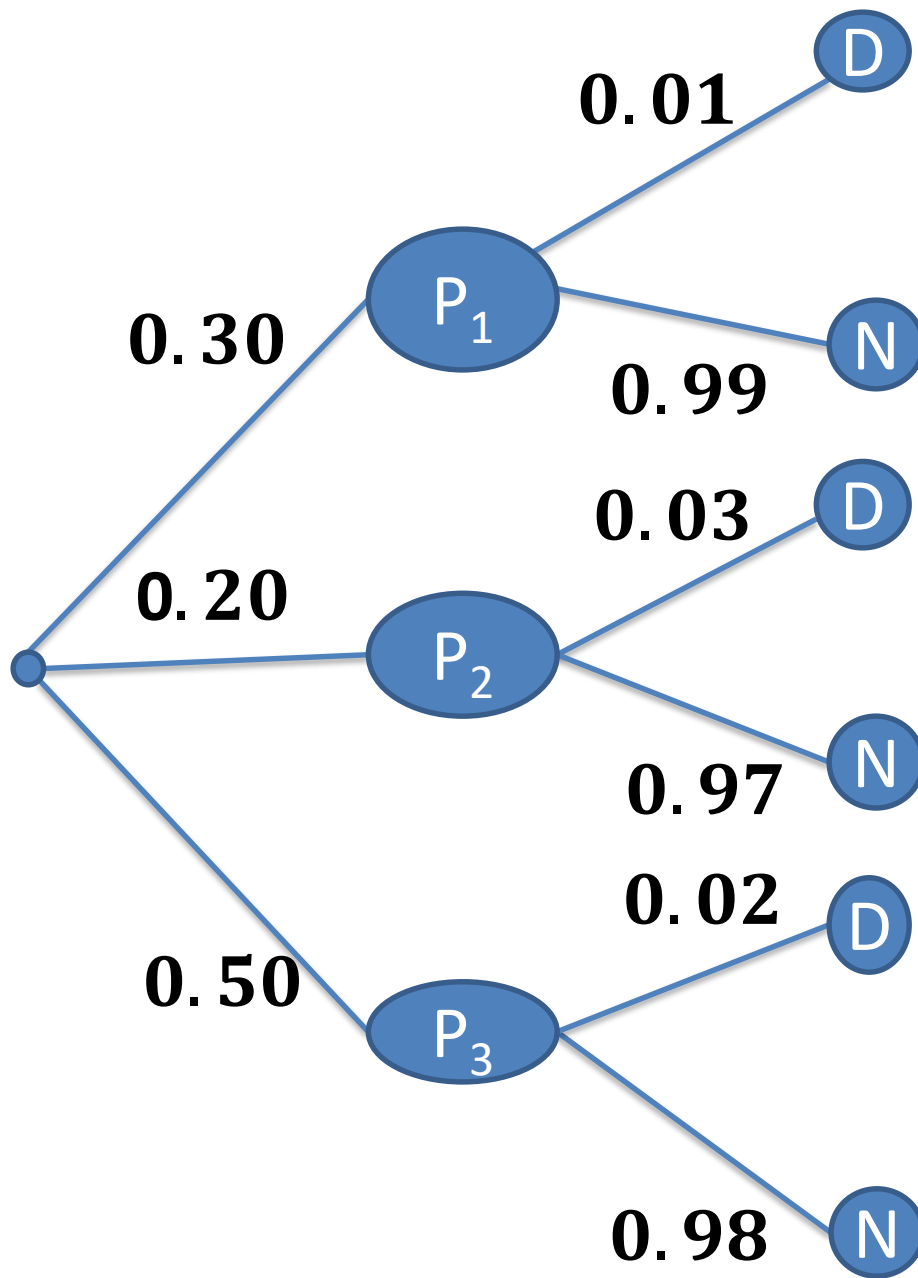
$$P(A | D) = \frac{P(A)P(D | A)}{P(A)P(D | A) + P(B)P(D | B) + P(C)P(D | C)}$$

$$\begin{aligned} P(A | D) &= \frac{(0.50)(0.03)}{(0.50)(0.03) + (0.30)(0.04) + (0.20)(0.05)} \\ &= \frac{15}{37} \text{ (or 0.4054)} \end{aligned}$$

Example: A manufacturing firm employs **three** analytical plans for the design and development of a particular product. For cost reasons, **all three** are used at varying times. In fact, plans **1**, **2**, and **3** are used for **30%**, **20%**, and **50%** of the products, respectively. The defect rate is different for the three procedures as follows:

$$P(D|P_1) = 0.01, P(D|P_2) = 0.03, P(D|P_3) = 0.02,$$

where $P(D|P_j)$ is the probability of a defective product, given **plan j** . If a random product was observed and found to be **defective**, which **plan** was **most likely** used and thus responsible?



Solution: Given

$P(P_1) = 0.30$, $P(P_2) = 0.20$, and $P(P_3) = 0.50$,

$P(D|P_1) = 0.01$, $P(D|P_2) = 0.03$, $P(D|P_3) = 0.02$

We have to find **$P(P_j|D)$ for $j = 1, 2, 3$.**

By Bayes' theorem,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

$$P(P_1|D) = \frac{P(P_1)P(D|P_1)}{P(P_1)P(D|P_1) + P(P_2)P(D|P_2) + P(P_3)P(D|P_3)}$$
$$= \frac{(0.30)(0.01)}{(0.30)(0.01) + (0.20)(0.03) + (0.50)(0.02)}$$

$$= \frac{0.003}{0.019}$$
$$= \mathbf{0.158}$$

$$\begin{aligned}
 P(P_2 | D) &= \frac{P(P_2)P(D | P_2)}{P(P_1)P(D | P_1) + P(P_2)P(D | P_2) + P(P_3)P(D | P_3)} \\
 &= \frac{(0.20)(0.03)}{(0.30)(0.01) + (0.20)(0.03) + (0.50)(0.02)} = \mathbf{0.316}
 \end{aligned}$$

$$\begin{aligned}
 P(P_3 | D) &= \frac{P(P_3)P(D | P_3)}{P(P_1)P(D | P_1) + P(P_2)P(D | P_2) + P(P_3)P(D | P_3)} \\
 &= \frac{(0.50)(0.02)}{(0.30)(0.01) + (0.20)(0.03) + (0.50)(0.02)} = \mathbf{0.526.}
 \end{aligned}$$

The conditional probability of a defect given **plan 3** is the largest of the **three**; thus a defective for a random product is most likely the result of the use of **plan 3**

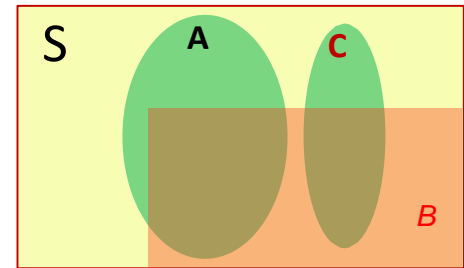
Conditional probabilities share properties of ordinary probabilities

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

I. $P(A | B) \geq 0$

II. $P(S | B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

III. $P(B | B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$



Conditional probabilities share properties of ordinary probabilities

If $A \cap C = \emptyset$ then $P(A \cup C \mid B) = P(A \mid B) + P(C \mid B)$

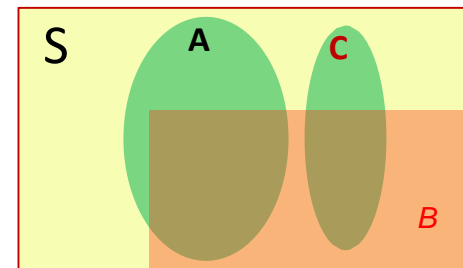
$$P(A \cup C \mid B) = \frac{P((A \cup C) \cap B)}{P(B)}$$

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

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

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

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



Conditional probabilities share properties of ordinary probabilities

Since conditional probabilities satisfy all of **the probability axioms**, any formula or theorem that we ever derive for ordinary probabilities will remain true for conditional probabilities as well.

Model based on conditional probabilities

- ❑ Let us now examine what conditional probabilities are good for.
- ❑ They are used to **revise a model** when we get **new information**, but there is another way in which they arise.
- ❑ We can use **conditional probabilities** to **build a multi-stage model** of a probabilistic experiment.

Model based on conditional probabilities

Example Radar Detection. If an **aircraft is present** in a certain area, a **radar detects** it and generates **an alarm** signal with **probability 0.99**. If an **aircraft is not present** the radar generates a **(false) alarm**, with probability **0.10**. We assume that an **aircraft is present** with probability **0.05**. What is the probability of **no aircraft presence** and a **false alarm**? What is the probability of **aircraft presence** and **no detection**?

Model based on conditional probabilities

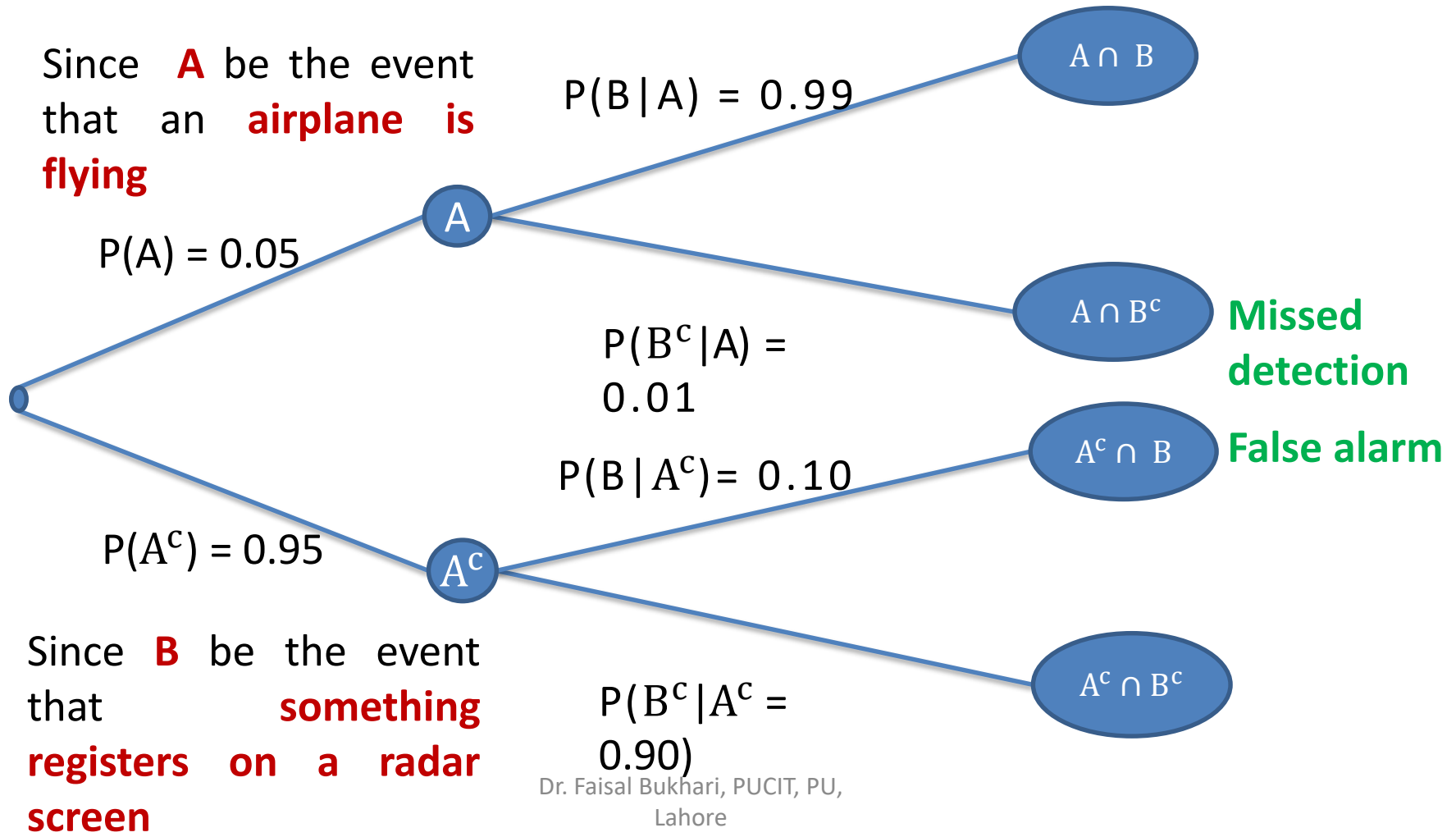
Let **A** be the event that an **airplane is flying**

Let **A^c** be the event that an **airplane is not flying**

Let **B** be the event that **something registers on a radar screen.**

Let **B^c** be the event that **something does not registers on a radar screen.**

Since **A** be the event that an **airplane is flying**



Dr. Faisal Bukhari, PUCIT, PU,
Lahore

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

$$\begin{aligned} P(A \cap B) &= P(A) \times P(B|A) \\ &= 0.05 \times 0.99 = \mathbf{0.0495 \text{ or } 4.95 \%} \end{aligned}$$

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) && \text{(total probability)} \\ &= P(A) \times P(B|A) + P(A^c) \times P(B|A^c) \\ &= 0.05 \times 0.99 + 0.95 \times 0.10 = \mathbf{0.1445 \text{ or } 14.45\%} \end{aligned}$$

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{0.0495}{0.1445} = \mathbf{0.3426 \text{ or } 34.26\%} \end{aligned}$$

$$\begin{aligned} \text{P(not present , false alarm)} &= \text{P}(A^c \cap B) \\ &= \text{P}(A^c)\text{P}(B | A^c) \\ &= 0.95 \times 0.10 \\ &= \mathbf{0.095 \text{ or } 9.5 \% \text{ ans}} \end{aligned}$$

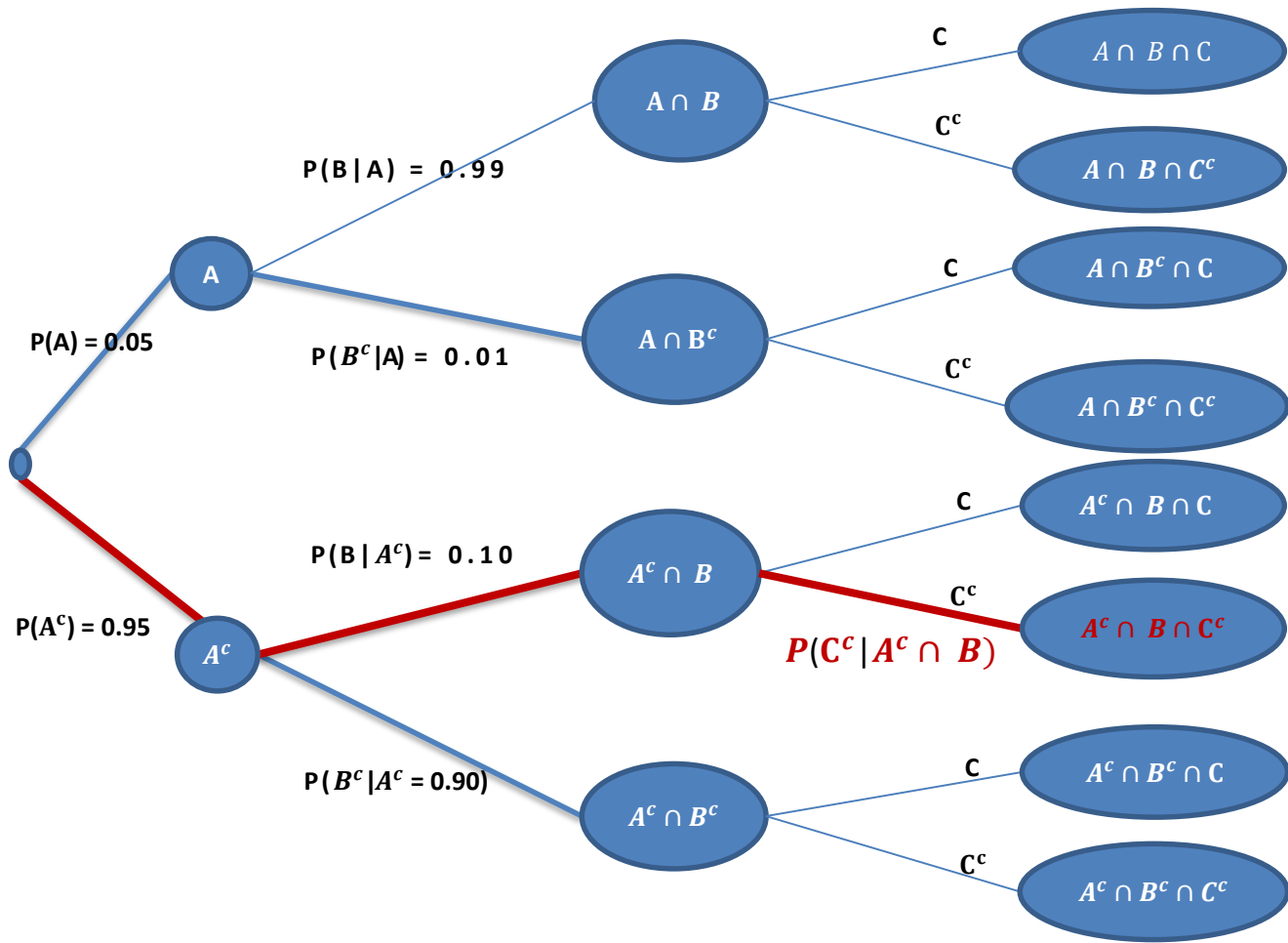
$$\begin{aligned} \text{P(present , no detection)} &= \text{P}(A \cap B^c) \\ &= \text{P}(A)\text{P}(B^c | A) \\ &= 0.05 \times 0.01 \\ &= \mathbf{0.0005 \text{ or } 0.05 \% \text{ ans}} \end{aligned}$$

The multiplication rule

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

$$\Rightarrow P(A \cap B) = P(A) \times P(B|A)$$

$$\Rightarrow P(A \cap B) = P(B) \times P(A|B)$$



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

$$\begin{aligned} P(A^c \cap B \cap C^c) &= P((A^c \cap B) \cap C^c) \\ &= P(A^c \cap B) \times P(C^c | A^c \cap B) \\ &= P(A^c) \times P(B | A^c) \times P(C^c | A^c \cap B) \end{aligned}$$

$$P(A_1 \cap A_2 \cap A_3 \dots \cap A_n) = P(A_1) \prod_{i=2}^n P(A_i | A_1 \cap A_2 \dots \cap A_{i-1})$$

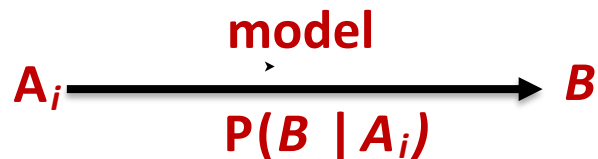
□ And this is the most general version of the **multiplication** rule and allows you to calculate the probability of several events happening by multiplying probabilities and conditional probabilities.

Bayes' rule and inference

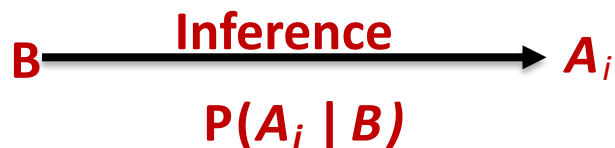
□ Bayesian inference

□ initial beliefs $P(A_i)$ on possible causes of an observed event B

□ model of the world under each A_i : $P(B | A_i)$



– draw conclusions about causes



Bayesian method

- ❑ Using **Bayes' rule**, a statistical methodology called the **Bayesian approach** has attracted a lot of attention in **applications**.
- ❑ An introduction to the **Bayesian method** will be discussed in coming lectures.