



Introduction to theory of probability and statistics

Lecture 3.

Independence and total probability rule

dr hab.inż. Katarzyna Zakrzewska, prof.AGH

Katedra Elektroniki, AGH

e-mail: zak@agh.edu.pl

<http://home.agh.edu.pl/~zak>

Outline:

- Independence of random events
- Total probability
- Problems using the material covered-
presentation of solutions
- Bayes' theorem

Multiplication rule

Definition of conditional probability:

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

can be rewritten to provide a general expression for the probability of the intersection of events:

$$P(A \cap B) = P(B / A)P(A) = P(A / B)P(B)$$

This formula is referred to as a **multiplication rule** for probabilities.

Example 3.1:

The probability that an automobile battery subject to high engine compartment temperature suffers low charging current is 0.7. The probability that a battery is subject to high engine compartment temperature is 0.05.

Solution to example 3.1:

Let C denote the event that a battery suffers low charging current, and let T denote the event that a battery is subject to high engine compartment temperature. The probability that a battery is subject to low charging current and high engine compartment temperature is:

$$P(C \cap T) = P(C | T)P(T) = 0.7 \times 0.05 = 0.035$$

Independence

From a definition of conditional probability

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

in a special case we get the following.

If $P(B|A) = P(B)$ than the outcome of A **has no influence on B**.

Events A and B can be treated as **independent**.

For two independent events we have:

$$P(A \cap B) = P(A) \cdot P(B)$$

Example 3.2:

Suppose a day's production of 850 manufactured parts contains 50 parts that do not meet customer requirements. Suppose two parts are selected from a batch, but the first part is replaced before the second part is selected (**sampling with replacement**). Calculate the probability of an event B , that the second part is defective given that the first part is defective (denoted as A). The probability needed can be expressed as $P(B|A)$.

Solution to example 3.2:

Because the first part is replaced prior to selecting the second part, the batch still contains 850 parts, of which 50 are defective. Therefore, the probability of B does not depend on whether or not the first part was defective. That is,

$$P(B | A) = P(B) = \frac{50}{850}$$

Also, the probability that both parts are defective is

$$P(A \cap B) = P(B | A) \cdot P(A) = \left(\frac{50}{850}\right) \cdot \left(\frac{50}{850}\right) = 0,0035$$

Independent events?

Example 3.3:

Similarly to example 3.2, a day's production of 850 manufactured parts contains 50 parts that do not meet customer requirements. Suppose two parts are selected from a batch, but the first part is **not replaced** before the second part is selected (**sampling without (w/o) replacement**).

Calculate the probability of an event B, that the second part is defective given that the first part is defective (denoted as A). The probability needed can be expressed as $P(B|A)$.

We suspect that these two events are not independent because the knowledge that the first part is defective suggests that it is less likely that the second part selected is defective.

But how to find the unconditional probability $P(B)$?

Independent events?

Solution to example 3.3:

Because of the specific type of sampling (w/o replacement), during the second attempt the set from which we draw elements contains 849 elements only, including 49 defective ones (the first element was defective). Thus the probability:

$$P(B | A) = \frac{49}{849}$$

Probability of event B depends on whether the first element was defective (event A) or not (event A'):

$$P(B | A') = \frac{50}{849}$$

In order to prove that A and B are not independent, we have to calculate the **unconditional probability P(B)**

Total probability rule

Multiplication rule:

$$P(A \cap B) = P(B / A)P(A) = P(A / B)P(B)$$

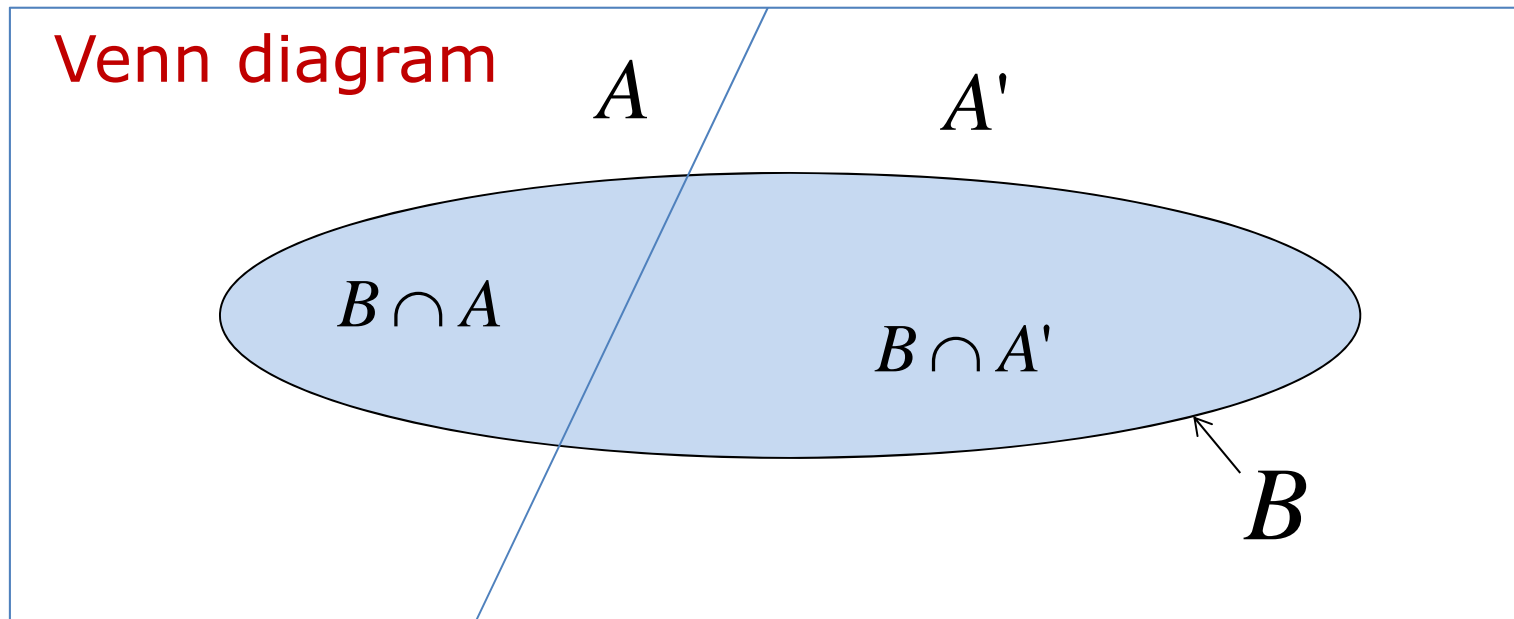
is useful for determining the probability of an event that depends on other events.

Example 3.4: Suppose that in semiconductor manufacturing the probability is 0.10 that a chip that is subjected to high levels of contamination during manufacturing causes a product failure. The probability is 0.005 that a chip that is not subjected to high contamination levels during manufacturing causes a product failure. In a particular production run, 20% of the chips are subject to high levels of contaminations. What is the probability that a product using one of these chips fails?

Total probability rule

For any event B : $B = (B \cap A) \cup (B \cap A')$

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A') = \\ &= P(B | A)P(A) + P(B | A')P(A') \end{aligned}$$



Total probability rule

Solution to problem 3.4: Let F denote the event that the product fails, and H denote the event that the chip is exposed to high levels of contamination. The information provided can be represented as

$$P(F | H) = 0.10 \quad \text{and} \quad P(F | H') = 0.005$$

$$P(H) = 0.20 \quad \text{and} \quad P(H') = 0.80$$

The requested probability $P(F)$ can be calculated as:

$$P(F) = P(F | H)P(H) + P(F | H')P(H')$$

$$P(F) = 0.10 (0.20) + 0.005 (0.80) = 0.024$$

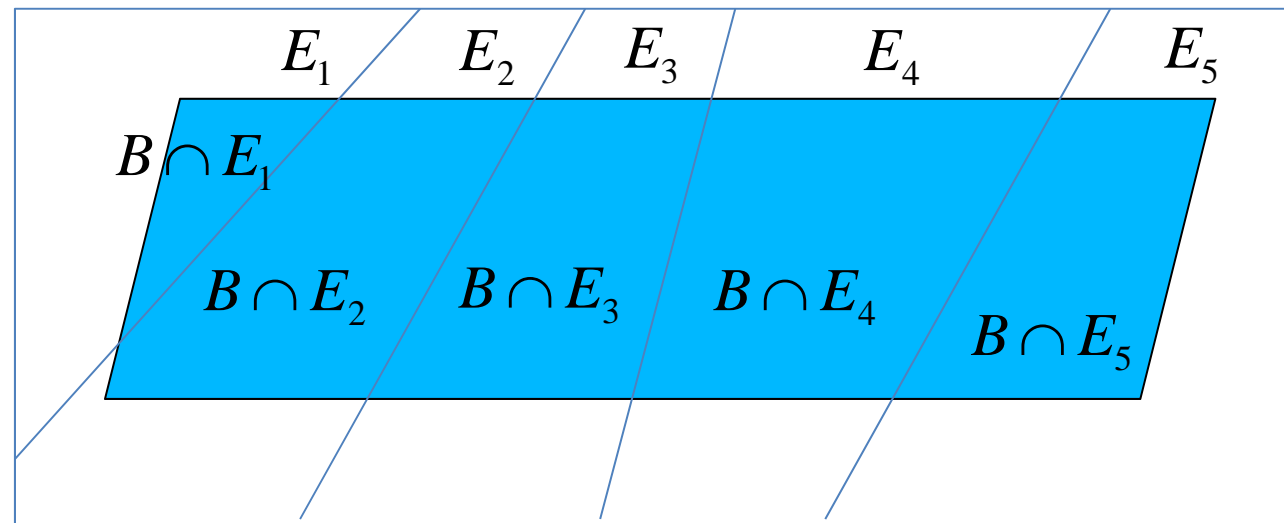
which can be interpreted as the weighted average of the two probabilities of failure.

Total probability rule (multiple events)

Assume E_1, E_2, \dots, E_k are k **mutually exclusive** and **exhaustive** sets. Then, probability of event B :

$$\begin{aligned}
 P(B) &= P(B \cap E_1) + P(B \cap E_2) + \dots + P(B \cap E_k) = \\
 &= P(B | E_1)P(E_1) + P(B | E_2)P(E_2) + \dots + P(B | E_k)P(E_k)
 \end{aligned}$$

where $\sum_{i=1}^k E_i = \Omega$



Independent events?

Solution to problem 3.3 (cont.): We have found that probability of event B that consists in drawing a defective element in a second attempt depends on the result of the first because it is carried out without replacement (event A – defective element in the first attempt; event A' – good element in the first attempt)

$$P(B|A) = \frac{49}{849} \qquad P(B|A') = \frac{50}{849}$$

Are A and B **independent**?

Finding the unconditional P(B) is somewhat difficult because the possible values of the first selection need to be considered

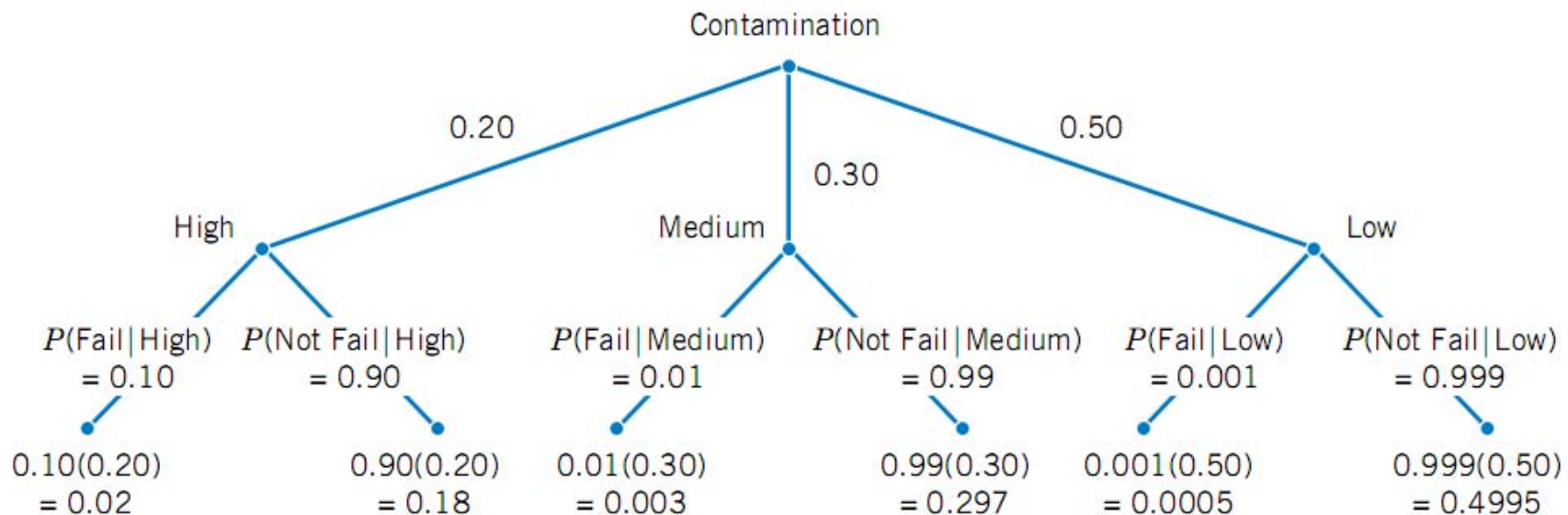
$$P(B) = P(B|A)P(A) + P(B|A')P(A')$$
$$P(B) = \left(\frac{49}{849}\right)\left(\frac{50}{850}\right) + \left(\frac{50}{849}\right)\left(\frac{800}{850}\right) = \frac{50}{850} \neq P(B|A)$$

Answer: events A and B **are not independent!**



Total probability rule – problem for individual work

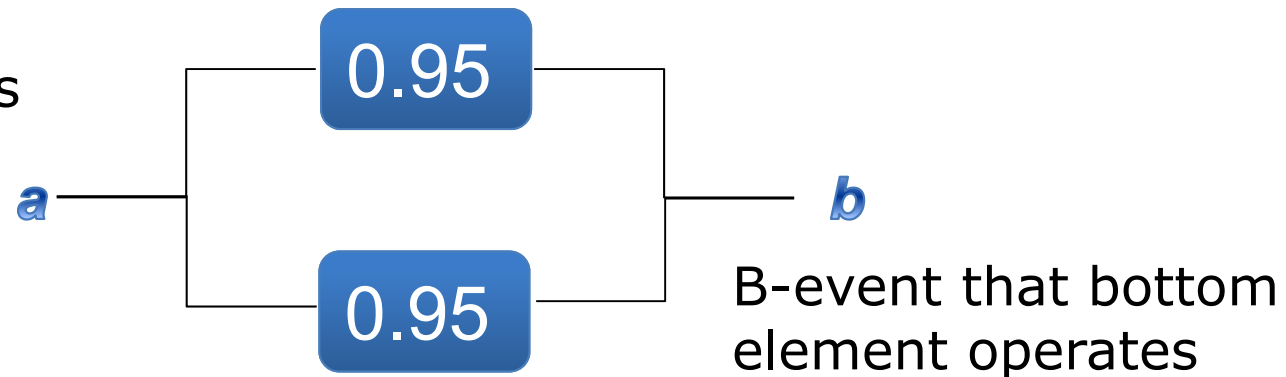
Probability of Failure	Level of contamination	Percentage of elements contaminated
0,1	High	20% (H)
0,01	Medium	30% (M)
0,001	Low	50% (L)



Problems to solve

Example 3.5: The following circuit operates only if there is a path of functional devices from left (a) to right (b). The probability that each device functions is shown on the graph. Assume that devices fail independently. What is a probability that the circuit operates?

T-event that top element operates



$$P(T \text{ or } B) = P(T \cup B) = 1 - P[(T \cup B)']$$

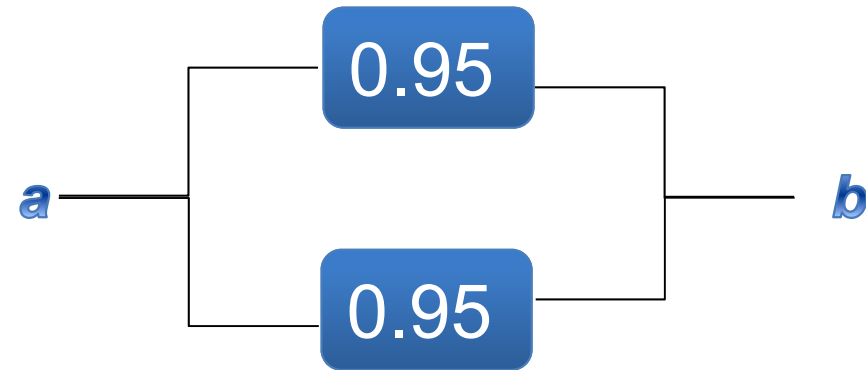
↑
probability of the complement of events

Problems to solve

Solution to problem 3.5: To solve this problem we have to apply DeMorgan's laws:

$$(T \cup B)' = T' \cap B'$$

$$(T \cap B)' = T' \cup B'$$



$$P(T \text{ or } B) = P(T \cup B) = 1 - P[(T \cup B)'] = 1 - P(T' \cap B')$$

Events B' and T' are independent, then:

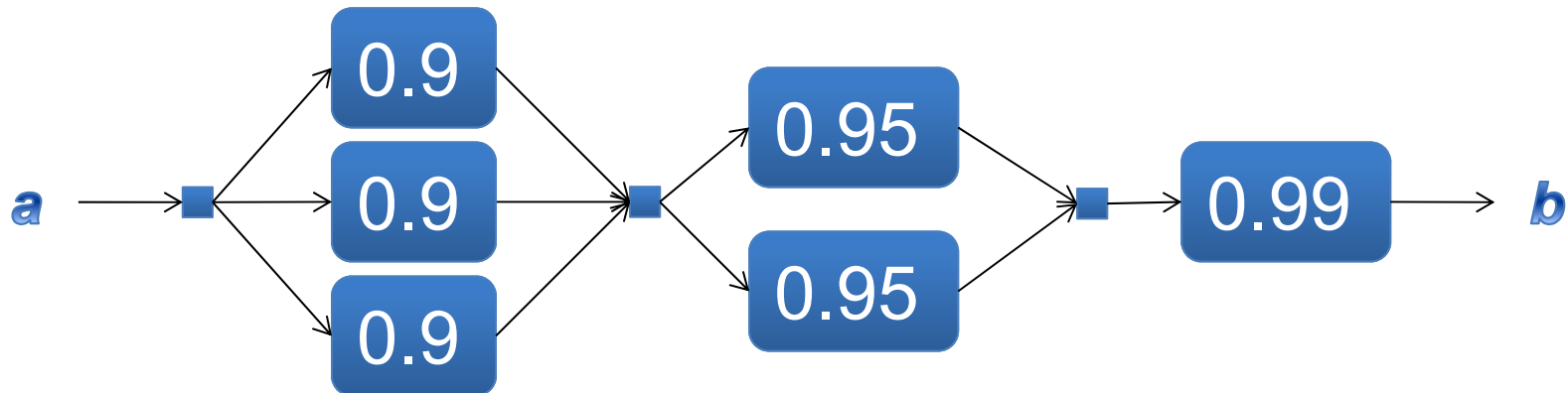
$$P(T' \text{ and } B') = P(T' \cap B') = P(T')P(B') = (1 - 0.95)^2 = 0.05^2$$

Answer: Probability that the circuit operates is:

$$P(T \text{ or } B) = 1 - P(T' \cap B') = 1 - 0.05^2 = 0.9975$$

Problems to solve at home

Problem 3.6: Calculate the probability that the circuit operates?



Problems to solve at home

Solution to 3.6:

In the first column:

$$1 - (1 - 0,9)^3 = 1 - (0,1)^3$$

Second column:

$$1 - (1 - 0,95)^2 = 1 - (0,05)^2$$

Probability that the circuit operates:

$$(1 - (0,1)^2)(1 - (0,05)^3)(0,99) = 0,987$$

Bayes' theorem

Simple transformation of a definition of the conditional probability leads to :

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

and then:

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

It could be useful in the following example (3.7) if the full table is unknown.

Problem 3.7

Event D denotes that a part is defective; F means that a thin film has a surface flaw

		Surface flaw:		
		Yes (F)	No (F')	Total
Defective	Yes (D)	10	28	38
	No (D')	30	332	362
Total		40	360	400

Data in table are complete and gives the number of thin films with surface flaws as well as those that are defective. **We can conclude that:**

$$P(D | F) = \frac{10}{40} = 0.25 \quad P(F) = \frac{40}{400} = 0.1 \quad P(D) = \frac{38}{400} = 0.095$$

Problem 3.7

When the table is complete then one can find directly:

$$P(F | D) = \frac{10}{38}$$

If not, $P(F/D)$ has to be calculated from Bayes' theorem as:

$$P(F | D) = \frac{P(D | F) \cdot P(F)}{P(D)}$$

$$P(F | D) = \frac{\frac{10}{40} \cdot \frac{40}{400}}{\frac{38}{400}} = \frac{10}{38}$$

Bayes' theorem

If E_1, E_2, \dots, E_k are k **mutually exclusive** and **exhaustive** events and B is any event,

$$P(E_1|B) = \frac{P(B|E_1)P(E_1)}{P(B|E_1)P(E_1) + P(B|E_2)P(E_2) + \dots + P(B|E_k)P(E_k)}$$

for $P(B) > 0$



Problem 3.8

Because a new medical procedure has been shown to be effective in the early detection of an illness, a medical screening of the population is proposed. The probability that the test correctly identifies someone with the illness as positive is $P(S/D)=0.99$, and the probability that the test correctly identifies someone without illness as negative is $P(S'/D')=0.95$. The incidence of the illness in the general population is $P(D)=0,0001$.

You take a test, and the result is positive. What is a probability that you have the illness?

Solution to problem 3.8: D-event that you are ill; S-event that the test signals positive; requested probability is $P(D/S)$.

The probability that the test correctly signals someone without illness as negative is $P(S'/D')=0.95$. Consequently, the probability of a positive test without illness $P(S/D')=1-P(S'/D')=0,05$

Problem 3.8

$$P(S/D') = 1 - P(S'/D') = 0,05$$

$$P(S/D) = 0,99; P(S'/D') = 0,95; P(D) = 0,0001$$

From Bayes' theorem

$$P(D/S) = \frac{P(S|D) \cdot P(D)}{P(S|D) \cdot P(D) + P(S|D') \cdot P(D')}$$

$$P(D/S) = \frac{0,99 \cdot 0,0001}{0,99 \cdot 0,0001 + 0,05 \cdot (1 - 0,0001)} = 0,002$$

Surprisingly, even though the test is effective, in the sense that $P(S/D)$ is high and $P(S'/D')$ is low, because the incidence of the illness in the general population is low, the chances are quite small that you actually have the disease even if the test is positive.