Lecture 1 - Introduction and basic definitions

Jan Bouda

FI MU

February 22, 2012

Part I

Motivation

What?

- Randomness
- Probability
- Statistics

Motivation of the course

- Probability is one of the central concepts of mathematics and also of computer science
- It allows us to model and study not only truly random processes, but also situation when we have incomplete information about the process.
- It is important when studying average behavior of algorithms, adversaries, ...

Applications of probability theory in computer science

- Information theory
- Coding theory
- Cryptography
- Random processes
- Randomized algorithms
- Complexity theory
- and many more ...

Bibliography


M. Mitzenmacher and E. Upfal
Probability and Computing
Cambridge University Press, 2005


G. Grimmett, D. Stirzaker
Probability and random processes
OUP Oxford, 2001


W. Feller.
An Introduction to Probability Theory and Its Applications
John Wiley & Sons, 1968


R. B. Ash and C. A. Doléans-Dade
Probability and Measure Theory
Harcourt Academic Press, 2000

Bibliography


T. M. Cover, J. A. Thomas.
Elements of Information Theory
John Wiley & Sons, 2006


D. R. Stinson.
Cryptography: Theory and Practice
Chapman & Hall/CRC, 2006


R. Motwani and P. Raghavan
Randomized Algorithms
Cambridge University Press, 2000

Goals of probability theory

- The main goal of the probability theory is to study random experiments.
- Random experiment is a model of physical or gedanken experiment, where we are uncertain about outcomes of the experiment, regardless whether the uncertainty is due to objective coincidences or our ignorance. We should be able to estimate how 'likely' respective outcomes are.
- Random experiment is specified by the set of possible outcomes and probabilities that each particular outcome occurs. Probability specifies how likely a particular outcome is to occur.

Examples of random experiments

A typical example of a random experiment is the coin tossing.

- Possible **outcomes** of this experiment are *head* and *tail*.
- Probability of each of these outcomes depends on physical properties of the coin, on the way it is thrown, dots
- In case of a fair coin (and fair coin toss) the probability to obtain head (tail) is $1/2$.
- In practice it is almost impossible to find unbiased coin, although theoretically we would expect a randomly chosen coin to be biased.
- Coin tossing is an important abstraction - analysis of the coin tossing can be applied to many other problems.

Examples of random experiments

Another example of a random experiment is throwing a six-sided die.

- Possible outcomes of a this experiment are symbols '1'–'6' representing respective facets of the die.
- Assuming the die is unbiased, the probability that a particular outcome occurs is the same for all outcomes and equals to $1/6$.

Examples of random experiments: Three balls in three cells

- In this experiment we have three cells and we sequentially and independently put each ball randomly into one of the cells.
- The atomic outcomes of this experiment correspond to positions of respective balls. Let us denote the balls as a, b and c.
- All atomic outcomes of this experiment are:

1. [abc][ ][ ]	10. [a][bc][ ]	19. [ ][a][bc]
2. [ ][abc][ ]	11. [b][a c][ ]	20. [ ][b][a c]
3. [ ][ ][abc]	12. [c][ab][ ]	21. [ ][c][ab]
4. [ab][c][ ]	13. [a][ ][bc]	22. [a][b][c]
5. [a c][b][ ]	14. [b][ ][a c]	23. [a][c][b]
6. [bc][a][ ]	15. [c][ ][ab]	24. [b][a][c]
7. [ab][ ][c]	16. [ ][ab][c]	25. [b][c][a]
8. [a c][ ][b]	17. [ ][a c][b]	26. [c][a][b]
9. [bc][ ][a]	18. [ ][bc][a]	27. [c][b][a]

Examples of random experiments

Example

Let us consider a source emitting 16 bit messages with uniform probability. This message is transmitted through a noisy channel that maps the first bit of the message to '0' and preserves all consecutive bits.
Possible outcomes of this experiment are all 16 bit messages, however, all messages starting with '1' have zero probability. On the other hand, all messages starting with '0' have the same probability 2^{-15}.

Examples of random experiments

Example

Let us consider a pushdown automaton A accepting language $L \subseteq \Sigma^*$. We choose randomly an n–symbol word $w \in \Sigma^n$ with probability $P(w)$ and pass it to the automaton A as an input. Let us suppose that the computation of A on w takes $\sharp_{A,w}$ steps. Then the average number of steps taken by the automaton A on n symbol input is

$$\sum_{w \in \Sigma^n} P(w) \sharp_{A,w}.$$

Part II

Sample space, events and probability

Random experiment

- The **(idealized) random experiment** is the central notion of the probability theory.
- Any random event will be modeled using an appropriate random experiment.
- Random experiment is specified (from mathematical point of view) by the set of possible outcomes and probabilities assigned to each of these outcomes.
- A single execution of a random experiment is called the **trial**.

Sample space

Definition

The set of the possible outcomes of a **random experiment** is called the **sample space** of the experiment and it will be denoted **S**. The outcomes of a random experiment (elements of the sample space) are denoted **sample points**.

- Every thinkable outcome of a random experiment is described by one, and only one, sample point.
- In this course we will concentrate on random experiments with countable sample space.

Sample space - examples

The sample space of the

- 'coin tossing' experiment is $\{head, tail\}$.
- 'throwing a six-sided die' experiment is $\{1, 2, 3, 4, 5, 6\}$.

Example

Let us consider the following experiment: we toss the coin until the 'head' appears. Possible outcomes of this experiment are

$$H, TH, TTH, TTTH, \ldots.$$

We may also consider the possibility that 'head' never occurs. In this case we have to introduce extra sample point denoted e.g. $\perp$.

Event

In addition to basic outcomes of a random experiment, we are often interested in more complicated events that represent a number of outcomes of a random experiment.
The event 'outcome is even' in the 'throwing a six-sided die' experiment corresponds to atomic outcomes '2', '4', '6'. Therefore, we represent an event of a random experiment as a subset of its sample space.

Event

Definition

An **event** of a random experiment with sample space **S** is any subset of **S**.

- The event $A \sim$'throwing by two independent dice results in sum 6' is $A = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$.
- Similarly, the event $B \sim$'two odd faces' is $B = \{(1,1),(1,3),\ldots(5,5)\}$.
- Every (atomic) outcome of an experiment is an event (single-element subset).
- Empty subset is an event.
- Set of all outcomes **S** is an event.

Event

Note that occurrence of a particular outcome may imply a number of events. In such a case these events occur simultaneously.

- The outcome $(3, 3)$ implies event 'the sum is 6' as well as the event 'two odd faces'.
- Events 'the sum is 6' and 'two odd faces' may occur simultaneously.
- Every compound event can be decomposed into atomic events (sample points), compound event is an **aggregate** of atomic events.

Example of events: Three balls in three cells

- The event A='there is more than one ball in one of the cells' corresponds to atomic outcomes 1-21. We say that the event A is an aggregate of events 1-21.
- The event B='first cell is not empty' is an aggregate of sample points 1,4-15,22-27.
- The event C is defined as 'both A and B occur'. It represents sample points 1,4-15.

1. [abc][ ][ ]	10. [a][bc][ ]	19. [ ][a][bc]
2. [ ][abc][ ]	11. [b][a c][ ]	20. [ ][b][a c]
3. [ ][ ][abc]	12. [c][ab][ ]	21. [ ][c][ab]
4. [ab][c][ ]	13. [a][ ][bc]	22. [a][b][c]
5. [a c][b][ ]	14. [b][ ][a c]	23. [a][c][b]
6. [bc][a][ ]	15. [c][ ][ab]	24. [b][a][c]
7. [ab][ ][c]	16. [ ][ab][c]	25. [b][c][a]
8. [a c][ ][b]	17. [ ][a c][b]	26. [c][a][b]
9. [bc][ ][a]	18. [ ][bc][a]	27. [c][b][a]

Example of events: Three balls in three cells

- You may observe that each of 27 sample points of this experiment occurs either in A or in B. Therefore the event 'either A or B or both occur' corresponds to the whole sample space and occurs with certainty.
- The event D='A does not occur' represents sample points 22-27 and can be rewritten as 'no cell remains empty'.
- The event 'first cell empty and no cell multiply occupied' is impossible since no sample point satisfies this condition.

1. [abc][ ][ ]	10. [a][bc][ ]	19. [ ][a][bc]
2. [ ][abc][ ]	11. [b][a c][ ]	20. [ ][b][a c]
3. [ ][ ][abc]	12. [c][ab][ ]	21. [ ][c][ab]
4. [ab][c][ ]	13. [a][ ][bc]	22. [a][b][c]
5. [a c][b][ ]	14. [b][ ][a c]	23. [a][c][b]
6. [bc][a][ ]	15. [c][ ][ab]	24. [b][a][c]
7. [ab][ ][c]	16. [ ][ab][c]	25. [b][c][a]
8. [a c][ ][b]	17. [ ][a c][b]	26. [c][a][b]
9. [bc][ ][a]	18. [ ][bc][a]	27. [c][b][a]

Algebra of events - notation

- The fact that a sample point $x \in \mathbf{S}$ is contained in event A is denoted $x \in \mathbf{A}$.
- The fact that two events A and B contain the same sample points is denoted $A = B$.
- We use $A = \emptyset$ to denote that event contains no sample points.
- To every event A there exists an event 'A does not occur'. It is denoted $\overline{A} \stackrel{def}{=} \{x \in \mathbf{S} | x \notin A\}$ and called the **complementary (negative) event**.
- The event 'both A and B occur' is denoted $A \cap B$.
- The event 'either A or B or both occur' is denoted $A \cup B$.
- The symbol $A \subseteq B$ signifies that every point of A is contained in B. We read it 'A implies B' and 'B is implied by A'.

Algebra of events - laws

E1 Commutative:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

E2 Associative:

$$A \cup (B \cup C) = (A \cup B) \cup C$$
$$A \cap (B \cap C) = (A \cap B) \cap C$$

E3 Distributive:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

E4 Identity:

$$A \cup \emptyset = A, \quad A \cap \mathbf{S} = A$$

E5 Complement:

$$A \cup \overline{A} = \mathbf{S}, \quad A \cap \overline{A} = \emptyset$$

Any relation valid in the algebra of events can be proved using these axioms.

Algebra of events - some relations

Using the previously introduced axioms we can derive e.g.

- Idempotent laws:
$$A \cup A = A, \quad A \cap A = A$$
- Domination laws:
$$A \cup \mathbf{S} = \mathbf{S}, \quad A \cap \emptyset = \emptyset$$
- Absorption laws:
$$A \cup (A \cap B) = A, \quad A \cap (A \cup B) = A$$
- de Morgan's laws:
$$\overline{(A \cup B)} = \overline{A} \cap \overline{B}, \quad \overline{(A \cap B)} = \overline{A} \cup \overline{B}$$
-
$$\overline{\overline{A}} = A$$
-
$$A \cup (\overline{A} \cap B) = A \cup B$$

Events

Definition

Events $A_1, A_2, \ldots, A_n$ are **mutually exclusive** if and only if

$$\forall i \neq jA_i \cap A_j = \emptyset.$$

Definition

Events $A_1, A_2, \ldots, A_n$ are **collectively exhaustive** if and only if

$$A_1 \cup A_2 \cdots \cup A_n = \mathbf{S}.$$

A set of events can be collectively exhaustive, mutually exclusive, both or neither. Mutually exclusive and collectively exhaustive list is called a **partition** of the sample space $\mathbf{S}$.

Example

Let us define the list of events $A_s = \{s\}$ for each sample point $s \in \mathbf{S}$. Such a list of events is a partition of $\mathbf{S}$.

Probability

- To complete our specification of the random experiment (as a mathematical model) we have to assign probabilities to sample points in the sample space.
- In many engineering application the so-called relative frequency is used interchangeably with probability. This is insufficient for our purposes, however, we may use the relative frequency as an approximation of probability in case we can perform a number of trials of the experiment, but we do not have theoretical description of a random experiment.

Definition

Let us suppose that we perform n trials of a random experiment and we obtain an outcome $s \in \mathbf{S}$ k times. Then the **relative frequency** of the outcome s is k/n.

Definition of Probability

By the symbol $P(s)$ we denote the probability of the sample point s. Analogously, we use the symbol $P(E)$ to denote the probability of an event E. The probability function must satisfy the Kolmogorov's axioms:

- A1 For any event A, $P(A) \geq 0$.
- A2 $P(\mathbf{S}) = 1$.
- A3 $P(A \cup B) = P(A) + P(B)$ provided that A and B are mutually exclusive events (i.e. $A \cap B = \emptyset$).

It is sufficient to assign probabilities to all sample points and we can calculate the probability of any event using these rules. Using the the axiom (A3) we can easily prove its generalized version for any finite number of events, however, for a countable infinite sequence of events we need a modified version of the axiom:

- A3' For any countable sequence of events $A_1, A_2, \ldots$ that are mutually exclusive it holds that

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Some relations

Using the Kolmogorov's axioms and axioms of the algebra of events we obtain

- Ra For any event A, $P(\overline{A}) = 1 - P(A)$.
- Rb If $\emptyset$ is the impossible event, $P(\emptyset) = 0$.
- Rc If A and B are (not necessarily exclusive) events then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Theorem (inclusion and exclusion)

If $A_1, A_2, \ldots, A_n$ are any events, then (see the principle of inclusion and exclusion of combinatorial mathematics)

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) =& P(A_1 \cup \cdots \cup A_n) \\ =& \sum_i P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) + \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) \\ & - \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \ldots A_n). \end{aligned}$$

Inclusion and exclusion - proof

inclusion and exclusion.

We use induction on the number of events n. For $n = 1$ we obtain the relation immediately. Let us suppose the statement holds for any union of $n-1$ events. Let us define the event $B = A_1 \cup \cdots \cup A_{n-1}$. Then

$$\bigcup_{i=1}^{n} A_i = B \cup A_n.$$

Using the result (Rc) we have

$$\begin{aligned} P\left(\bigcup_{i-1}^{n} A_i\right) &= P(B \cup A_n) \\ &= P(B) + P(A_n) - P(B \cap A_n). \end{aligned} \tag{1}$$

Inclusion and exclusion - proof continued

inclusion and exclusion.

Using the distributivity of intersection and union we have

$$B \cap A_n = (A_1 \cap A_n) \cup \cdots \cup (A_{n-1} \cap A_n)$$

is a union of $n-1$ events and thus we can apply the inductive hypothesis to obtain

$$\begin{aligned} P(B \cap A_n) = &\sum_{i=1}^{n-1} P(A_i \cap A_n) - \sum_{i<j;i,j=1}^{n-1} P[(A_i \cap A_n) \cap (A_j \cap A_n)] + \cdots + \\ &+ (-1)^{n-2} P[(A_1 \cap A_n) \cap (A_2 \cap A_n) \cap \cdots \cap (A_{n-1} \cap A_n)] = \\ = &\sum_{i=1}^{n-1} P(A_i \cap A_n) - \sum_{i<j;i,j=1}^{n-1} P(A_i \cap A_j \cap A_n) + \cdots + \\ &+ (-1)^{n-2} P(A_1 \cap A_2 \cap \cdots \cap A_{n-1} \cap A_n) \end{aligned} \tag{2}$$

Inclusion and exclusion - proof continued

inclusion and exclusion.

Also, since B is a union of $n-1$ events, the inductive hypothesis gives

$$P(B) = \sum_{i=1}^{n-1} P(A_i) - \sum_{i<j;i,j=1}^{n-1} P(A_i \cap A_j) + \cdots + \\ + (-1)^{n-2} P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \tag{3}$$

Substituting (2) and (3) into (1) gives the desired result. □

The probability space

Definition

Let $\mathbf{S}$ be a countable set. Then $\mathcal{F} \subseteq 2^{\mathbf{S}}$ is a σ-**field** of subsets of $\mathbf{S}$ if and only if it is closed under countable unions and complement.

Definition

A **probability space** is a triple $(\mathbf{S}, \mathcal{F}, P)$, where $\mathbf{S}$ is a set, $\mathcal{F}$ is a σ-field of subsets of $\mathbf{S}$ and P is a probability on $\mathcal{F}$ satisfying axioms (A1)-(A3').

In this formal definition $\mathbf{S}$ plays the role of the sample space, $\mathcal{F}$ is the set of events on $\mathbf{S}$ and P is the function assigning probability to each event.

Designing a random experiment

The design of a random experiment in a real situation is not of interest for the theory of probability, however, it is one of the basic skills good computer scientists need. Usually it is practical to follow this procedure:

- Identify the sample space - set of mutually exclusive and collectively exhaustive events. It is advised to choose the elements in the way that they cannot be further subdivided. You can always define aggregate events.
- Assign probabilities to elements in **S**. This assignment must be consistent with the axioms (A1)-(A3). Probabilities are usually either result of a careful theoretical analysis, or based on estimates obtained from past experience.
- Identify events of interest - they are usually described by statements and should be reformulated in terms of subsets of **S**.
- Compute desired probabilities - calculate the probabilities of interesting events using the axioms (A1)-(A3').

Balls and cells revisited

Example

Let us consider random placement of r balls into n cells. This generalization of the original (three balls and three cells) experiment is treated in an analogous manner, except that the number of sample points increases rapidly with r and n. In example, for $r = 4$ and $n = 3$ we have 81 points, for $r = n = 10$ there are 10^{10} sample points.

Balls and cells revisited

We described the experiments in terms of holes and balls, but this experiment can be equivalently applied to a number of practical situations. The only difference is the verbal description.

- Birthdays: The possible configuration of birthdays of r people corresponds to the random placement of r balls into $n = 365$ cells (assuming every year has 365 days).
- Elevator: Elevator starts with r passengers and stops in n floors.
- Dice: A throw with r dice corresponds to placing r balls into 6 holes. In case of a coin tosses we have $n = 2$ holes.
- Exam from IV111: The exam from Probability in Computer Science corresponds to placement of 37 (the number of students) balls into $n = 6$ holes (A,B,C,D,E,F).

Balls and cells revisited

Let us return to the first example with three balls and three cells and suppose that the balls are not distinguishable, implying e.g. that we do not distinguish atomic events 4, 5 and 6. Atomic events in the new experiment (placing three indistinguishable balls into three cells) are

```
1.[***][   ][   ]   6.[ * ][** ][   ]
2.[   ][***][   ]   7.[ * ][   ][** ]
3.[   ][   ][***]   8.[   ][** ][ * ]
4.[** ][ * ][   ]   9.[   ][ * ][** ]
5.[** ][   ][ * ]  10.[ * ][ * ][ * ]
```

- It is irrelevant for our theory whether the real balls are indistinguishable or not.
- Even if they are we may decide to treat them as indistinguishable, it is often even preferable.
- Dice may be colored to make them distinguishable, but it depends purely on our decision whether we use this possibility or not.

Balls and cells revisited

Another example is the exam from Probability in Computer Science.

- Each student is distinguishable, but to judge statistical outcomes of the exam, such as probability distribution of marks, it is useless to complicate the experiment by distinguishing respective students.
- The sample points of experiment with undistinguishable balls correspond to aggregates of experiment with distinguishable balls. In example, the atomic event 4. corresponds to aggregate event of sample points 4-6 in the original experiment.

The concrete situation dictates this choice. Our theory begins after the proper model has been chosen.

Part III

Conditional probability, independent events and Bayes' rule

Conditional probability

- Let us suppose that a random experiment was executed, but we did not learn the outcome. The only information we are given is that a particular event B occured.
- Our goal is to determine the probability of event A, however, the information that B occured may change our prediction of probability that A occured.
- We want to compute the conditional probability of the event A given that the event B occurred, shortly the conditional probability of A given B.

Conditional probability

Given that B occurred we know that the outcome of the experiment $o \in B$ and $o \notin \overline{B}$. For every atomic outcome s we derive

$$P(s|B) = \begin{cases} \frac{P(s)}{P(B)} & \text{if } s \in B, \\ 0 & s \in \overline{B}. \end{cases}$$

In this way the probabilities assigned to points in B are scaled up by $1/P(B)$. We obtain

$$\sum_{s \in B} P(s|B) = 1.$$

B is our 'sample space' now.

Conditional probability

The conditional probability of any event can be obtained by summing probabilities of its sample points, using $A = (A \cap B) \cup (A \cap \overline{B})$ we have

$$\begin{aligned} P(A|B) &\overset{def}{=} \sum_{s \in A} P(s|B) \\ &= \sum_{s \in A \cap B} P(s|B) + \sum_{s \in A \cap \overline{B}} P(s|B) \\ &= \sum_{s \in A \cap B} P(s|B) \\ &= \sum_{s \in A \cap B} \frac{P(s)}{P(B)} \\ &= \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0. \end{aligned}$$

Conditional probability

Definition

The **conditional probability** of A given B is

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

if $P(B) \neq 0$ and is undefined otherwise.

In this way we directly obtain

$$P(A \cap B) = \begin{cases} P(B)P(A|B) & \text{if } P(B) \neq 0, \\ P(A)P(B|A) & \text{if } P(A) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Conditional probability

Example

Four (distinguishable) balls are placed successively into four cells with all 4^4 outcomes being equally probable. Given that the first two balls are in different cells (event B), what is the probability that one cell contains exactly three balls (event A)? The number of sample points in events is $|B| = 12 \cdot 4^2$ and $|A \cap B| = 4 \cdot 3 \cdot 2$ and therefore we obtain that $P(A|B) = 2/16$ in contrast to $P(A) = 3/16$.

Independence of events

If the probability of the event A does not change regardless of whether event B occurred, i.e. $P(A|B) = P(A)$, we conclude that events A and B are independent.
Using the definition of the conditional probability for $P(A) \neq 0 \neq P(B)$ we have

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

We obtain the standard definition of independents of events

Definition

Events A and B are said to be **independent** if

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

Independence of events - remarks

- This relation is symmetric in A and B as desired - when A is independent of B, then also B is independent of A.
- If A and B are mutually exclusive, then $P(A \cap B) = 0$. If they are also independent, we obtain either $P(A) = 0$ or $P(B) = 0$.
- If an event A is independent on itself, due to re-normalization we obtain that either $P(A) = 0$ or $P(A) = 1$. Equivalently, $P(A) = P(A \cap A) = P(A)P(A) = [P(A)]^2$.
- A and B are independent and B and C are independent does not imply that A and C are independent. This relation is not transitive.
- If A and B are independent, then also A and $\overline{B}$, $\overline{A}$ and B, and $\overline{A}$ and $\overline{B}$ are independent.

Independence of events

Definition

Events $A_1, A_2, \ldots A_n$ are **(mutually) independent** if and only if for any set $i_1, i_2, \ldots i_k \in \{1 \ldots n\}$ $(2 \leq k \leq n)$ of distinct indices it holds that

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k}) \tag{4}$$

Provided that the Eq. (4) holds, we can replace any occurrence of A_i in this expression by $\overline{A}_i$ and the new expression will be valid too.
Compare the properties of $P(A_1 \cap A_2 \cdots \cap A_n)$ when $A_1, A_2, \ldots A_n$ are mutually independent and when they are mutually exclusive.
If the events $A_1, A_2, \ldots A_n$ are such that every pair is independent, then such events are called **pairwise independent**. It does not imply that events are mutually independent.

Independence of events-example

Example

Consider the experiment with throwing two dice. Let the sample space be $\mathbf{S} = \{(i,j)|1 \leq i,j \leq 6\}$. We assign to each outcome the probability $1/36$, i.e. we assume the uniform probability distribution. Let us define the events

A="first throw results in 1, 2 or 3."
B="first throw results in 3, 4 or 5."
C="the sum of both faces is 9."

Then $A \cap B = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$, $A \cap C = \{(3,6)\}$, $B \cap C = \{(3,6),(4,5),(5,4)\}$, and $A \cap B \cap C = \{(3,6)\}$. We obtain the probabilities

$$\begin{aligned} P(A \cap B) &= 1/6 \neq P(A)P(B) = 1/4 \\ P(A \cap C) &= 1/36 \neq P(A)P(C) = 1/18 \\ P(B \cap C) &= 1/12 \neq P(B)P(C) = 1/18. \end{aligned}$$

On the other hand, $P(A \cap B \cap C) = 1/36 = P(A)P(B)P(C)$.

Bayes' rule

- Suppose that Events B and $\overline{B}$ partition the sample space **S**.
- We can define $S' = \{B, \overline{B}\}$ and using the probabilities $P(B)$ and $P(\overline{B})$ the set S' has properties similar to sample space, except that there is many-to-one correspondence between outcomes of the experiment and elements of S'.
- Such a space is called **event space**.
- In general, any list of mutually exclusive and collectively exhaustive events forms an event space.

Theorem (of total probability)

Let A be an event and $S' = \{B_1, \ldots B_n\}$. Then

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Bayes' rule

The theorem of total probability is useful in the situation when we have information that the event A occurred and we are interested in conditional probabilities $P(B_i|A)$. We obtain the **Bayes' rule**

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_i P(A|B_i)P(B_i)}.$$

$P(B_i|A)$ is called the **a posteriori probability**.

Part IV

Repeated experiment execution - Bernoulli trials

Bernoulli trials

- Let us consider a random experiment with two possible outcomes, e.g. 'success' and 'failure'.
- we assign to the outcomes probabilities p and q, respectively. $p + q = 1$.
- Consider a compound experiment which is composed of a sequence of n independent executions (trials) of this experiment.
- Such a sequence is called a **sequence of Bernoulli trials**.

Example

Consider the binary symmetric channel, i.e. model of a noisy channel which transmits bit faithfully with probability p and inverts it with probability $q = 1 - p$. Probability of error on each transmitted bit is independent.

Bernoulli trials

Let 0 denotes 'failure' and 1 'success'. Let S_n be a sample space of an experiment involving n Bernoulli trials defined as

$$\begin{aligned} S_1 &= \{0, 1\} \\ S_2 &= \{(0,0), (0,1), (1,0), (1,1)\} \\ S_n &= \{2^n \text{ ordered } n\text{-tuples of 0s and 1s}\} \end{aligned}$$

We have the probability assignment over S_1: $P(0) = q \geq 0, P(1) = p \geq 0$, and $p + q = 1$. We would like to derive probabilities for S_n. Let A_i be the event 'success on trial i' and $\overline{A}_i$ denote 'failure on trial i'. Then $P(A_i) = p$ and $P(\overline{A}_i) = q$. Let us consider an element $s \in S_n$ such that

$$s = (\overbrace{1, 1, \ldots, 1}^{k}, \overbrace{0, 0, \ldots, 0}^{n-k}).$$

Bernoulli trials

The elementary event $\{s\}$ can be expressed as

$$\{s\} = A_1 \cap A_2 \cap \cdots \cap A_k \cap \overline{A}_{k+1} \cap \cdots \cap \overline{A}_n$$

with the probability

$$\begin{aligned} P(\{s\}) =& P(A_1 \cap A_2 \cap \cdots \cap A_k \cap \overline{A}_{k+1} \cap \cdots \cap \overline{A}_n) \\ =& P(A_1)P(A_2)\ldots P(A_k)P(\overline{A}_{k+1})\ldots P(\overline{A}_n) \end{aligned}$$

from independence. Finally,

$$P(\{s\}) = p^k q^{n-k}.$$

Similarly, to any other sample point with k 1s and $n-k$ 0s the same probability is assigned.