# Applied Probability and Stochastic Processes

Second Edition

Frank Beichelt

CRC Press
Taylor & Francis Group
A Chapman & Hall Book

# Applied Probability and Stochastic Processes

Second Edition

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# APPLIED PROBABILITY AND STOCHASTIC PROCESSES

## SECOND EDITION


FRANK BEICHELT

UNIVERSITY OF THE WITWATERSRAND

JOHANNESBURG, SOUTH AFRICA



CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742



Printed on acid-free paper
Version Date: 20160421

International Standard Book Number-13: 978-1-4822-5764-9 (Hardback)



**Visit the Taylor & Francis Web site at**
**http://www.taylorandfrancis.com**

**and the CRC Press Web site at**
**http://www.crcpress.com**

# CONTENTS

## 3 MULTIDIMENSIONAL RANDOM VARIABLES



## 4 FUNCTIONS OF RANDOM VARIABLES

## 5 INEQUALITIES AND LIMIT THEOREMS



# PART II STOCHASTIC PROCESSES

## 6 BASICS OF STOCHASTIC PROCESSES



## 7 RANDOM POINT PROCESSES

## 8 DISCRETE-TIME MARKOV CHAINS



## 9 CONTINUOUS-TIME MARKOV CHAINS

## 10 MARTINGALES

## 11 BROWNIAN MOTION



## 12 SPECTRAL ANALYSIS OF STATIONARY PROCESSES

# PREFACE TO THE SECOND EDITION

The book is a self-contained introduction into elementary probability theory and stochastic processes with special emphasis on their applications in science, engineering, finance, computer science and operations research. It provides theoretical foundations for modeling time-dependent random phenomena in these areas and illustrates their application through the analysis of numerous, practically relevant examples. As a non-measure theoretic text, the material is presented in a comprehensible, application-oriented way. Its study only assumes a mathematical maturity which students of applied sciences acquire during their undergraduate studies in mathematics. The study of stochastic processes and its fundament, probability theory, as of any other mathematically based science, requires less routine effort, but more creative work on one's own. Therefore, numerous exercises have been added to enable readers to assess to which extent they have grasped the subject. Solutions to many of the exercises can be downloaded from the website of the Publishers or the exercises are given together with their solutions. A complete solutions manual is available to instructors from the Publishers. To make the book attractive to theoretically interested readers as well, some important proofs and challenging examples and exercises have been included. 'Starred' exercises belong to this category. The chapters are organized in such a way that reading a chapter usually requires knowledge of some of the previous ones. The book has been developed in part as a course text for undergraduates and for self-study by non-statisticians. Some sections may also serve as a basis for preparing senior undergraduate courses.

The text is a thoroughly revised and supplemented version of the first edition so that it is to a large extent a new book: The part on probability theory has been completely rewritten and more than doubled. Several new sections have been included in the part about stochastic processes as well: Time series analysis, random walks, branching processes, and spectral analysis of stationary stochastic processes. Theoretically more challenging sections have been deleted and mainly replaced with a comprehensive numerical discussion of examples. All in all, the volume of the book has increased by about a third.

This book does not extensively deal with data analysis aspects in probability and stochastic processes. But sometimes connections between probabilistic concepts and the corresponding statistical approaches are established to facilitate the understanding. The author has no doubt the book will help students to pass their exams and practicians to apply stochastic modeling in their own fields of expertise.

The author is thankful for the constructive feedback from many readers of the first edition. Helpful comments to the second edition are very welcome as well and should be directed to: Frank.Beichelt@wits.ac.za.

Johannesburg, March 2016 *Frank Beichelt*

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

## SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| □ ■ ● | symbols after an example, a theorem, a definition |
| $f(t) \equiv c$ | $f(t) = c$ <u>for all</u> $t$ being element of the domain of definition of $f$ |
| $f * g$ | convolution of two functions $f$ and $g$ |
| $f^{*(n)}$ | $n$th convolution power of $f$ |
| $\hat{f}(s),\ L\{f, s\}$ | Laplace transform of a function $f$ |
| $o(x)$ | Landau order symbol |
| $\delta_{ij}$ | Kronecker symbol |

**Probability Theory**

| | |
|---|---|
| $X,\ Y,\ Z$ | random variables |
| $E(X),\ Var(X)$ | mean (expected) value of $X$, variance of $X$ |
| $f_X(x),\ F_X(x)$ | probability density function, (cumulative probability) distribution function of $X$ |
| $F_Y(y\|x),\ f_Y(y\|x)$ | conditional distribution function, density of $Y$ given $X = x$ |
| $X_t,\ F_t(x)$ | residual lifetime of a system of age $t$, distribution function of $X_t$ |
| $E(Y\|x)$ | conditional mean value of $Y$ given $X = x$ |
| $\lambda(x),\ \Lambda(x)$ | failure rate, integrated failure rate (hazard function) |
| $N(\mu, \sigma^2)$ | normally distributed random variable (normal distribution) with mean value $\mu$ and variance $\sigma^2$ |
| $\varphi(x),\ \Phi(x)$ | probability density function, distribution function of a standard normal random variable $N(0, 1)$ |
| $f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$ | joint probability density function of $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ |
| $F_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$ | joint distribution function of $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ |
| $Cov(X, Y),\ \rho(X, Y)$ | covariance, correlation coefficient between $X$ and $Y$ |
| $M(z)$ | $z$-transform (moment generating function) of a discrete random variable or of its probability distribution, respectively |

**Stochastic Processes**

| | |
|---|---|
| $\{X(t),\ t \in \mathbf{T}\},\ \{X_t,\ t \in \mathbf{T}\}$ | continuous-time, discrete-time stochastic process with parameter space $\mathbf{T}$ |
| $\mathbf{Z}$ | state space of a stochastic process |
| $f_t(x),\ F_t(x)$ | probability density, distribution function of $X(t)$ |
| $f_{t_1, t_2, \ldots, t_n}(x_1, x_2, \ldots, x_n),\ F_{t_1, t_2, \ldots, t_n}(x_1, x_2, \ldots, x_n)$ | joint density, distribution function of $(X(t_1), X(t_2), \ldots, X(t_n))$ |
| $m(t)$ | trend function of a stochastic process |
| $C(s,t)$ | covariance function of a stochastic process |
| $C(\tau)$ | covariance function of a stationary stochastic process |

| | |
|---|---|
| $C(t),\ \{C(t),\ t \ge 0\}$ | compound random variable, compound stochastic process |
| $\rho(s,t)$ | correlation function of a stochastic process |
| $\{T_1, T_2, ...\}$ | random point process |
| $\{Y_1, Y_2, ...\}$ | sequence of interarrival times, renewal process |
| $N$ | integer-valued random variable, discrete stopping time |
| $\{N(t),\ t \ge 0\}$ | (random) counting process |
| $N(s,t)$ | increment of a counting process in $(s, t]$ |
| $H(t),\ H_1(t)$ | renewal function of an ordinary, delayed renewal processs |
| $A(t)$ | forward recurrence time, point availability |
| $B(t)$ | backward recurrence time |
| $R(t),\ \{R(t),\ t \ge 0\}$ | risk reserve, risk reserve process |
| $A,\ A(t)$ | stationary (long-run) availability, point availability |
| $p_{ij},\ p_{ij}^{(n)}$ | one-step, $n$-step transition probabilities of a homogeneous, discrete-time Markov chain |
| $p_{ij}(t);\ q_{ij},\ q_i$ | transition probabilities; conditional, unconditional transition rates of a homogeneous, continuous-time Markov chain |
| $\{\pi_i;\ i \in \mathbf{Z}\}$ | stationary state distribution of a homogeneous Markov chain |
| $\pi_0$ | extinction probability, vacant probability (sections 8.5, 9.7) |
| $\lambda_j,\ \mu_j$ | birth, death rates |
| $\lambda,\ \mu,\ \rho$ | arrival rate, service rate, traffic intensity $\lambda/\mu$ (in queueing models) |
| $\mu_i$ | mean sojourn time of a semi-Markov process in state $i$ |
| $\mu$ | drift parameter of a Brownian motion process with drift |
| $W$ | waiting time in a queueing system |
| $L$ | lifetime, cycle length, queue length, continuous stopping time |
| $L(x)$ | first-passage time with regard to level $x$ |
| $L(a,b)$ | first-passage time with regard to level $\min(a, b)$ |
| $\{B(t),\ t \ge 0\}$ | Brownian motion (process) |
| $\sigma^2,\ \sigma$ | $\sigma^2 = Var(B(1))$ variance parameter, volatility |
| $\{S(t),\ t \ge 0\}$ | seasonal component of a time series (section 6.4), standardized Brownian motion (chapter 11). |
| $\{\bar{B}(t),\ 0 \le t \le 1\}$ | Brownian bridge |
| $\{D(t),\ t \ge 0\}$ | Brownian motion with drift |
| $M(t)$ | absolute maximum of the Brownian motion (with drift) in $[0, t]$ |
| $M$ | absolute maximum of the Brownian motion (with drift) in $[0, \infty)$ |
| $\{U(t),\ t \ge 0\}$ | Ornstein-Uhlenbeck process, integrated Brownian motion process |
| $\omega,\ w$ | circular frequency, bandwidth |
| $s(\omega),\ S(\omega)$ | spectral density, spectral function (chapter 12) |

# Introduction

Is the world a well-ordered entirety,
or a random mixture,
which nevertheless is called world-order?

*Marc Aurel*

Random influences or phenomena occur everywhere in nature and social life. Their consideration is an indispensable requirement for being successful in natural, economical, social, and engineering sciences. Random influences partially or fully contribute to the variability of parameters like wind velocity, rainfall intensity, electromagnetic noise levels, fluctuations of share prices, failure time points of technical units, timely occurrences of births and deaths in biological populations, of earthquakes, or of arrivals of customers at service centers. Random influences induce *random events*. An event is called *random* if on given conditions it can occur or not. For instance, the events that during a thunderstorm a certain house will be struck by lightning, a child will reach adulthood, at least one shooting star appears in a specified time interval, a production process comes to a standstill for lack of material, a cancer patient survives chemotherapy by 5 years are random. Border cases of random events are the *deterministic events*, namely the *certain event* and the *impossible event*. On given conditions, a deterministic (impossible) event will always (never) occur. For instance, it is absolutely sure that lead, when heated to a temperature of over $327.5\,^0C$ will become liquid, but that lead during the heating process will turn to gold is an impossible event. Random is the shape, liquid lead assumes if poured on an even steel plate, and random is also the occurrence of events which are predicted from the form of these castings to the future. Even if the reader is not a lottery, card, or dice player, she/he will be confronted in her/his daily routine with random influences and must take into account their implications: When your old coffee machine fails after an unpredictable number of days, you go to the supermarket and pick a new one from the machines of your favorite brand. At home, when trying to make your first cup of coffee, you realize that you belong to the few unlucky ones who picked by chance a faulty machine. A car driver, when estimating the length of the trip to his destination, has to take into account that his vehicle may start only with delay, that a traffic jam could slow down the progress, and that scarce parking opportunities may cause further delay. Also, at the end of a year the overwhelming majority of the car drivers realize that having taken out a policy has only enriched the insurance company. Nevertheless, they will renew their policy because people tend to prefer moderate regular cost, even if they arise long-term, to the risk of larger unscheduled cost. Hence it is not surprising that insurance companies belonged to the first institutions that had a direct practical interest in making use of methods for the quantitative evaluation of random influences and gave in turn important impulses for the development of such methods. It is the probability theory, which provides the necessary mathematical tools for their work.

*Probability theory deals with the investigation of regularities random events are subjected to.*

The existence of such *statistical* or *stochastic regularities* may come as a surprise to philosophically less educated readers, since at first glance it seems to be paradoxical to combine regularity and randomness. But even without philosophy and without probability theory, some simple regularities can already be illustrated at this stage:

1) When throwing a fair die once, then one of the integers from 1 to 6 will appear and no regularity can be observed. But if a die is thrown repeatedly, then the fraction of throws with outcome 1, say, will tend to 1/6, and with increasing number of throws this fraction will converge to the value 1/6. (A die is called *fair* if each integer has the same chance to appear.)
2) If a specific atom of a radioactive substance is observed, then the time from the beginning of its observation to its disintegration cannot be predicted with certainty, i.e., this time is random. On the other hand, one knows the *half-life period* of a radioactive substance, i.e., one can predict with absolute certainty after which time from say originally 10 gram (trillions of atoms) of the substance exactly 5 gram is left.
3) Random influences can also take effect by superimposing purely deterministic processes. A simple example is the measurement of a physical parameter, e.g., the temperature. There is nothing random about this parameter when it refers to a specific location at a specific time. However, when this parameter has to be measured with sufficiently high accuracy, then, even under always the same measurement conditions, different measurements will usually show different values. This is, e.g., due to the degree of inaccuracy, which is inherent to every measuring method, and to subjective moments. A statistical regularity in this situation is that with increasing number of measurements, which are carried out independently and are not biased by systematic errors, the arithmetic mean of these measurements converges towards the true temperature.
4) Consider the movement of a tiny particle in a container filled with a liquid. It moves along zig-zag paths in an apparently chaotic motion. This motion is generated by the huge number of impacts the particle is exposed to with surrounding molecules of the fluid. Under average conditions, there are about $10^{21}$ collisions per second between particle and molecules. Hence, a deterministic approach to modeling the motion of particles in a fluid is impossible. This movement has to be dealt with as a random phenomenon. But the pressure within the container generated by the vast number of impacts of fluid molecules with the sidewalls of the container is constant.

Examples 1 to 4 show the nature of a large class of statistical regularities:

*The superposition of a large number of random influences leads under certain conditions to deterministic phenomena.*

Deterministic regularities (law of falling bodies, spreading of waves, Ohm's law, chemical reactions, theorem of Pythagoras) can be verified in a single experiment if the underlying assumptions are fulfilled. But, although statistical regularities can be proved in a mathematically exact way just as the theorem of Pythagoras or the rules of differentiation and integration of real functions, their experimental verification requires a huge number of repetitions of one and the same experiment. Even leading scientists spared no expense to do just this. The *Comte de Buffon* (1707 – 1788) and the mathematician *Karl Pearson* (1857 – 1936) had flipped a fair coin several thousand times and recorded how often 'head' had appeared. The following table shows their results ($n$ number of total flippings, $m$ number of outcome 'head'):

| Scientist | $n$ | $m$ | $m/n$ |
|---|---|---|---|
| Buffon | 4040 | 2048 | 0.5080 |
| Pearson | 12000 | 6019 | 0.5016 |
| Pearson | 24000 | 12012 | 0.5005 |

Thus, the more frequently a coin is flipped, the more approaches the ratio $m/n$ the value 1/2 (compare with example 1 above). In view of the large number of flippings, this principal observation is surely not a random result, but can be confirmed by all those readers who take pleasure in repeating these experiments. However, nowadays the experiment 'flipping a coin' many thousand times is done by a computer with a 'virtual coin' in a few seconds. The ratio $m/n$ is called the *relative frequency* of the occurrence of the random event 'head appears.'

Already the expositions made so far may have convinced many readers that random phenomena are not figments of human imagination, but that their existence is objective reality. There have been attempts to deny the existence of random phenomena by arguing that if all factors and circumstances, which influence the occurrence of an event are known, then an absolutely sure prediction of its occurrence is possible. In other words, the protagonists of this thesis consider the creation of the concept of randomness only as a sign of 'human imperfection.' The young *Pierre Simeon Laplace* (1729 – 1827) believed that the world is down to the last detail governed by deterministic laws. Two of his famous statements concerning this are: 'The curve described by a simple molecule of air in any gas is regulated in a manner as certain as the planetary orbits. The only difference between them lies in our ignorance.' And: 'Give me all the necessary data, and I will tell you the exact position of a ball on a billiard table' (after having been pushed). However, this view has proved futile both from the philosophical and the practical point of view. Consider, for instance, a biologist who is interested in the movement of animals in the wilderness. How on earth is he supposed to be in a position to collect all that information, which would allow him to predict the movements of only one animal in a given time interval with absolute accuracy? Or imagine the amount of information you need and the corresponding software to determine the exact path of a particle, which travels in a fluid, when there are $10^{21}$ collisions with surrounding molecules per second. It is an unrealistic and impossible task to deal with problems like that in a deterministic way. The physicist *Marian von Smoluchowski* (1872 – 1917) wrote in a paper published in 1918 that 'all theories are inadequate, which consider randomness as an unknown partial cause of an event. The chance of the occurrence of an event can only depend on the conditions, which have influence on the event, but not on the degree of our knowledge.'

Already at a very early stage of dealing with random phenomena the need arose to quantify the *chance*, the *degree of certainty*, or the *likelihood* for the occurrence of random events. This had been done by defining the probability of random events and by developing methods for its calculation. For now the following explanation is given: The probability of a random event is a number between 0 and 1. The impossible event has probability 0, and the certain event has probability 1. The probability of a random event is the closer to 1, the more frequently it occurs. Thus, if in a long series of experiments a random event $A$ occurs more frequently than a random event $B$, then $A$ has a larger probability than $B$. In this way, assigning probabilities to random events allows comparisons with regard to the frequency of their occurrence under identical conditions. There are other approaches to the definition of probability than the classical (frequency) approach, to which this explanation refers. For beginners the frequency approach is likely the most comprehensible one.

Gamblers, in particular dice gamblers, were likely the first people, who were in need of methods for comparing the chances of the occurrence of random events, i.e., the chances of winning or losing. Already in the medieval poem *De Vetula* of *Richard de Fournival* (ca 1200–1250) one can find a detailed discussion about the total number of possibilities to achieve a certain number, when throwing 3 dice. *Geronimo Cardano* (1501 – 1576) determined in his book *Liber de Ludo Aleae* the number of possibilities to achieve the total outomes 2, 3, ..,12, when two dice are thrown. For instance, there are two possibilities to achieve the outcome 3, namely (1,2) and (2,1), whereas 2 will be only then achieved, when (1,1) occurs. (The notation $(i, j)$ means that one die shows an $i$ and the other one a $j$.) *Galileo Galilei* (1564 – 1642) proved by analogous reasoning that, when throwing 3 dice, the probability to get the (total) outcome 10 is larger than the probability to get a 9. The gamblers knew this from their experience, and they had asked Galilei to find a mathematical proof. The *Chevalier de Méré* formulated three problems related to games of chance and asked the French mathematician *Blaise Pascal* (1623 – 1662) for solutions:

1) What is more likely, to obtain at least one 6 when throwing a die four times, or in a series of 24 throwings of two dice to obtain at least once the outcome (6,6)?

2) How many time does one have to throw two dice at least so that the probability to achieve the outcome (6,6) is larger than 1/2?

3) In a game of chance, two equivalent gamblers need each a certain number of points to become winners. How is the stake to fairly divide between the gamblers, when for some reason or other the game has to be prematurely broken off ? (This problem of the *fair division* had been already formulated before *de Méré*, e.g., in the *De Vetula*.)

Pascal sent these problems to *Pierre Fermat* (1601 – 1665) and both found their solutions, although by applying different methods. It is generally accepted that this work of *Pascal* and *Fermat* marked the beginning of the development of probability theory as a mathematical discipline. Their work has been continued by famous scientists as *Christian de Huygens* (1629 – 1695), *Jakob Bernoulli* (1654 – 1705), *Abraham de Moivre* (1667 – 1754), *Carl Friedrich Gauss* (1777 – 1855), and last but not least by *Simeon Denis de Poisson* (1781 – 1840). However, probability theory was out of its infancy only in the thirties of the twentieth century, when the Russian mathematician *Andrej Nikolajewič Kolmogorov* (1903 – 1987) found the solution of one of the famous Hilbert problems, namely to put probability theory as any other mathematical discipline on an axiomatic foundation.

Nowadays, probability theory together with its applications in science, medicine, engineering, economy et al. are integrated in the field of *stochastics.* The linguistic origin of this term can be found in the Greek word *stochastikon.* (Originally, this term denoted the ability of seers to be correct with their forecasts.) Apart from probability theory, *mathematical statistics* is the most important part of stochastics. A key subject of it is to infer by probabilistic methods from a sample taken from a set of interesting objects, called among else *sample space* or *universe,* to parameters or properties of the sample space (*inferential statistics*). Let us assume we have a lot of 10 000 electronic units. To obtain information on what percentage of these units is faulty, we take a sample of 100 units from this lot. In the sample, 4 units are faulty. Of course, this figure does not imply that there are exactly 400 faulty units in the lot. But inferential statistics will enable us to construct lower and upper bounds for the percentage of faulty units in the lot, which limit the 'true percentage' with a given high probability. Problems like this led to the development of an important part of mathematical statistics, the *statistical quality control*. Phenomena, which depend both on random and deterministic influences, gave rise to the theory of *stochastic processes*. For instance, meteorological parameters like temperature and air pressure are random, but obviously also depend on time and altitude. Fluctuations of share prices are governed by chance, but are also driven by periods of economic up and down turns. Electromagnetic noise caused by the sun is random, but also depends on the periodical variation of the intensity of sunspots.

*Stochastic modeling in operations research* comprises disciplines like queueing theory, reliability theory, inventory theory, and decision theory. All of them play an important role in applications, but also have given many impulses for the theoretical enhancement of the field of stochastics. *Queueing theory* provides the theoretical fundament for the quantitative evaluation and optimization of queueing systems, i.e., service systems like workshops, supermarkets, computer networks, filling stations, car parks, and junctions, but also military defense systems for 'serving' the enemy. *Inventory theory* helps with designing warehouses (storerooms) so that they can on the one hand meet the demand for goods with sufficiently high probability, and on the other hand keep the costs for storage as small as possible. The key problem with dimensioning queueing systems and storage capacities is that flows of customers, service times, demands, and delivery times of goods after ordering are subject to random influences. A main problem of *reliability theory* is the calculation of the reliability (survival probability, availability) of a system from the reliabilities of its subsystems or components. Another important subject of reliability theory is modelling the aging behavior of technical systems, which incidentally provides tools for the survival analysis of human beings and other living beings. Chess automats got their intelligence from the *game theory*, which arose from the abstraction of games of chance. But opponents within this theory can also be competing economic blocs or military enemies. Modern communication would be impossible without *information theory.* This theory provides the mathematical foundations for a reliable transmission of information although signals may be subject to noise at the transmitter, during transmission, and at the receiver. In order to verify stochastic regularities, nowadays no scientist needs to manually repeat thousands of experiments. Computers do this job much more efficiently. They are in a position to virtually replicate the operation of even highly complex systems, which are subjected to random influences, to any degree of accuracy. This process is called (*Monte Carlo*) *simulation.* More and very fruitful applications of stochastic (probabilistic) methods exist in fields like physics (kinetic gas theory, thermodynamics, quantum theory), astronomy (stellar statistics), biology (genetics, genomics, population dynamic), artificial intelligence (inference under undertainty), medicine, genomics, agronomy and forestry (design of experiments, yield prediction) as well as in economics (time series analysis) and social sciences. There is no doubt that probabilistic methods will open more and more possibilities for applications, which in turn will lead to a further enhancement of the field of stochastics.

More than 300 hundreds years ago, the famous Swiss mathematician *Jakob Bernoulli* proposed in his book *Ars Conjectandi* the recognition of *stochastics* as an independent new science, the subject of which he introduced as follows:

*To conjecture about something is to measure its probability: The Art of conjecturing or the Stochastic Art is therefore defined as the art of measuring as exactly as possible the probability of things so that in our judgement and actions we always can choose or follow that which seems to be better, more satisfactory, safer and more considered.*

In line with Bernoulli's proposal, an independent science of stochastics would have to be characterized by two features:

1) The subject of stochastics is uncertainty caused by randomness and/or ignorance.

2) Its methods, concepts, and language are based on mathematics.

But even now, in the twenty-first century, an independent science of stochastics is still far away from being officially established. There is, however, a powerful support for such a move by internationally leading academics; see *von Collani* (2003).

# PART I

# Probability Theory

There is no credibility in sciences in which no mathematical theory can be applied, and no credibility in fields which have no connections to mathematics.

*Leonardo da Vinci*

# CHAPTER 1

## Random Events and Their Probabilities

### 1.1 RANDOM EXPERIMENTS

If water is heated up to $100^0 C$ at an air pressure of 101 325 $Pa$, then it will inevitably start boiling. A motionless pendulum, when being pushed, will start swinging. If ferric sulfate is mixed with hydrochloric acid, then a chemical reaction starts, which releases hydrogen sulfide. These are examples for experiments with deterministic outcomes. Under specified conditions they yield an outcome, which had been known in advance.

Somewhat more complicated is the situation with *random experiments* or *experiments with random outcome.* They are characterized by two properties:

1. Repetitions of the experiment, even if carried out under identical conditions, generally have different outcomes.

2. The possible outcomes of the experiment are known.

Thus, the outcome of a random experiment cannot be predicted with certainty. This implies that the study of random experiments makes sense only if they can be repeated sufficiently frequently under identical conditions. Only in this case *stochastic* or *statistical regularities* can be found.

Let $\mathbf{\Omega}$ be the set of possible outcomes of a random experiment. This set is called *sample space, space of elementary events,* or *universe.* Examples of random experiments and their respective sample spaces are:

1) Counting the number of traffic accidents a day in a specified area: $\mathbf{\Omega} = \{0, 1, ...\}$.

2) Counting the number of cars in a parking area with maximally 200 parking bays at a fixed time point: $\mathbf{\Omega} = \{0, 1, ..., 200\}$.

3) Counting the number of shooting stars during a fixed time interval: $\mathbf{\Omega} = \{0, 1, ...\}$.

4) Recording the daily maximum wind velocity at a fixed location: $\mathbf{\Omega} = [0, \infty)$.

5) Recording the lifetimes technical systems or organisms: $\mathbf{\Omega} = [0, \infty)$.

6) Determining the number of faulty parts in a set of 1000: $\mathbf{\Omega} = \{0, 1, ..., 1000\}$.

7) Recording the daily maximum fluctuation of a share price: $\mathbf{\Omega} = [0, \infty)$.

8) The total profit sombody makes with her/his financial investments a year.
This 'profit' can be negative, i.e. any real number can be the outcome: $\mathbf{\Omega} = (-\infty, +\infty)$.

9) Predicting the outcome of a wood reserve inventory in a forest stand: $\mathbf{\Omega} = [0, \infty)$.

10) a) Number of eggs a sea turtle will bury at the beach: $\mathbf{\Omega} = \{0, 1, ...\}$.

b) Will a baby turtle, hatched from such an egg, reach the water? $\mathbf{\Omega} = \{0, 1\}$ with meaning 0: no, 1: yes.

As the examples show, in the context of a random experiment, the term 'experiment' has a more general meaning than in the customary sense.

A random experiment may also contain a deterministic component. For instance, the measurement of a physical quantity should ideally yield the exact (deterministic) parameter value. But in view of random measurement errors and other (subjective) influences, this ideal case does not materialize. Depending on the degree of accuracy required, different measurements, even if done under identical conditions, may yield different values of one and the same parameter (length, temperature, pressure, amperage,...).

## 1.2 RANDOM EVENTS

A possible outcome $\omega$ of a random experiment, i.e. any $\omega \in \mathbf{\Omega}$, is called an *elementary event* or a *simple event.*

1) The sample space of the random experiment 'throwing two dice consists of 36 simple elements: $\mathbf{\Omega} = \{(i,j),\ i,j = 1, 2, \cdots, 6\}$. The gambler wins if the sum $i+j$ is at least 10. Hence, the 'winning simple events' are $(5,5)$, $(5,6)$, $(6,5)$, and $(6,6)$.

2) In a delivery of 100 parts some may be defective. A subset (sample) of $n = 12$ parts is taken, and the number $N$ of defective parts in the sample is counted. The elementary events are 0,1,...,12 (possible numbers of defective parts in the sample). The delivery is rejected if $N \geq 4$.

3) In training, a hunter shoots at a cardboard dummy. Given that he never fails the dummy, the latter is the sample space $\mathbf{\Omega}$, and any possible impact mark at the dummy is an elementary event. Crucial subsets to be hit are e.g. 'head' or 'heart.'

Already these three examples illustrate that often not single elementary events are interesting, but sets of elementary events. Hence it is not surprising that concepts and results from set theory play a key role in formally establishing probability theory. For this reason, next the reader will be reminded of some basic concepts of set theory.

**Basic Concepts and Notation from Set Theory** A set is given by its *elements.* We can consider the set of all real numbers, the set of all rational numbers, the set of all people attending a performance, the set of buffalos in a national park, and so on. A set is called *discrete* if it is a finite or a *countably infinite* set. By definition, a countably infinite set can be written as a sequence. In other words, its elements can be numbered. If a set is infinite, but not countably infinite, then it is called *nondenumerable.* Nondenumerable sets are for instance the whole real axis, the positive half-axis, a finite subinterval of the real axis, or a geometric object (area of a circle, target).

Let $A$ and $B$ be two sets. In what follows we assume that all sets $A, B, \ldots$ considered are subsets of a 'universal set' $\mathbf{\Omega}$. Hence, for any set $A$, $A \subseteq \mathbf{\Omega}$.

$A$ is called a *subset* of $B$ if each element of $A$ is also an element of $B$.

Symbol: $A \subseteq B$.

The *complement* of $B$ with regard to $A$ contains all those elements of $B$ which are not element of $A$.

Symbol: $B \backslash A$

In particular, $\overline{A} = \mathbf{\Omega} \backslash A$ contains all those elements which are not element of $A$.

The *intersection* of $A$ and $B$ contains all those elements which belong both to $A$ and $B$.

Symbol: $A \cap B$

The *union* of $A$ and $B$ contains all those elements which belong to $A$ or $B$ (or to both).

Symbol: $A \cup B$

These relations between two sets are illustrated in Figure 1.1 (Venn diagram). The whole shaded area is $A \cup B$.

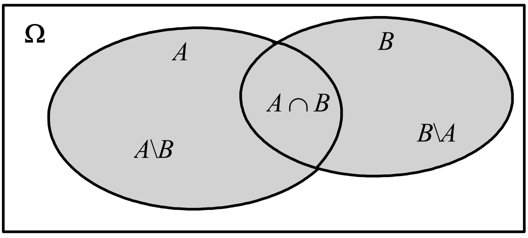



Figure 1.1 Venn diagram

For any sequence of sets $A_1, A_2, \cdots, A_n$, intersection and union are defined as

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n\,, \qquad \bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n\,.$$

*De Morgan Rules for* 2 *Sets*

$$\overline{A \cup B} = \overline{A} \cap \overline{B}\,, \quad \overline{A \cap B} = \overline{A} \cup \overline{B}\,. \tag{1.1}$$

*De Morgan Rules for n Sets*

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A}_i, \quad \overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A}_i\,. \tag{1.2}$$

**Random Events** A *random event* (briefly: *event*) $A$ is a subset of the set $\mathbf{\Omega}$ of all possible outcomes of a random experiment, i.e. $A \subseteq \mathbf{\Omega}$.

> *A random event A is said to have* occurred *as a result of a random experiment if the observed outcome* $\omega$ *of this experiment is an element of A:* $\omega \in A$.

The empty set $\varnothing$ is the *impossible event* since, for not containing any elementary event, it can never occur. Likewise, $\mathbf{\Omega}$ is the *certain event,* since it comprises all possible outcomes of the random experiment. Thus, there is nothing random about the events $\varnothing$ and $\mathbf{\Omega}$. They are actually *deterministic events.* Even before having completed a random experiment, we are absolutely sure that $\mathbf{\Omega}$ will occur and $\varnothing$ will not.

Let $A$ and $B$ be two events. Then the set-theoretic operations introduced above can be interpreted in terms of the occurrence of random events as follows:

$A \cap B$ is the event that both $A$ and $B$ occur,

$A \cup B$ is the event that $A$ or $B$ (or both) occur,

If $A \subseteq B$ ($A$ is a subset of $B$), then the occurrence of $A$ implies the occurrence of $B$.

$A \backslash B$ is the set of all those elementary events which are elements of $A$, but not of $B$. Thus, $A \backslash B$ is the event that $A$ occurs, but not $B$. Note that (see Figure 1.1)

$$A \backslash B = A \backslash (A \cap B). \tag{1.3}$$

The event $\overline{A} = \mathbf{\Omega} \backslash A$ is called the *complement of A*. It consists of all those elementary events, which are not in $A$.

Two events $A$ and $B$ are called *disjoint* or (*mutually*) *exclusive* if their joint occurrence is impossible, i.e. if $A \cap B = \varnothing$. In this case the occurrence of $A$ implies that $B$ cannot occur and vice versa. In particular, $A$ and $\overline{A}$ are disjoint for any event $A \subseteq \mathbf{\Omega}$.

***Short Terminology***

| | |
|---|---|
| $A \cap B$ | $A$ and $B$ |
| $A \cup B$ | $A$ or $B$ |
| $A \subseteq B$ | $A$ implies $B$, $B$ follows from $A$ |
| $A \backslash B$ | $A$ but not $B$ |
| $\overline{A}$ | $A$ not |

**Example 1.1** Let us consider the random experiment 'throwing a die' with sample space $\mathbf{\Omega} = \{1, 2, \cdots, 6\}$ and the random events $A = \{2, 3\}$ and $B = \{3, 4, 6\}$. Then, $A \cap B = \{3\}$ and $A \cup B = \{2, 3, 4, 6\}$. Thus, if a 3 had been thrown, then both the events $A$ and $B$ have occurred. Hence, $A$ and $B$ are not disjoint. Moreover, $A \backslash B = \{2\}$, $B \backslash A = \{4, 6\}$, and $\overline{A} = \{1, 4, 5, 6\}$. □

**Example 1.2** Two dice $D_1$ and $D_2$ are thrown. The sample space is

$$\mathbf{\Omega} = \{(i_1, i_2),\ i_1, i_2 = 1, 2, \cdots, 6\}.$$

Thus, an elementary event ω consists of two integers indicating the results $i_1$ and $i_2$ of $D_1$ and $D_2$, respectively. Let $A = \{i_1 + i_2 \le 3\}$ and $B = \{i_1 / i_2 = 2\}$. Then,

$$A = \{(1, 1),\ (1, 2),\ (2, 1)\}, \quad B = \{(2, 1),\ (4, 2),\ (6, 3)\}.$$

Hence,

$$A \cap B = \{(2, 1\}\}, \quad A \cup B = \{(1, 1),\ (1, 2),\ (2, 1), (4, 2),\ (6, 3)\}$$

and
$$A \backslash B = \{(1, 1),\ (1, 2)\}.$$
□

**Example 1.3** A company is provided with power by three generators $G_1$, $G_2$, and $G_3$. The company has sufficient power to maintain its production if only two out of the three generators are operating. Let $A_i$ be the event that generator $G_i$, $i = 1, 2, 3$, is operating, and $B$ be the event that at least two generators are operating. Then,

$$B = A_1 A_2 A_3 \cup A_1 A_2 \overline{A_3} \cup A_1 \overline{A_2} A_3 \cup \overline{A_1} A_2 A_3.$$
□

## 1.3 PROBABILITY

The aim of this section consists in constructing rules for determining the probabilities of random events. Such a rule is principally given by a function $P$ on the set $\boldsymbol{E}$ of all random events $A$: $P = P(A)$, $A \in \boldsymbol{E}$.

Note that in this context $A$ is an element of the set $\boldsymbol{E}$ so that the notation $A \subseteq \boldsymbol{E}$ would not be correct. Moreover, not all subsets of $\mathbf{\Omega}$ need to be random events, i.e., the set $\boldsymbol{E}$ need not necessarily be the set of all possible subsets of $\mathbf{\Omega}$.

The function $P$ assigns to every event $A$ a number $P(A)$, which is its probability. Of course, the construction of such a function cannot be done arbitrarily. It has to be done in such a way that some obvious properties are fulfilled. For instance, if $A$ implies the occurrence of the event $B$, i.e. $A \subseteq B$, the $B$ occurs more frequently than $A$ so that the relation $P(A) \le P(B)$ should be valid. If in addition the function $P$ has properties $P(\varnothing) = 0$ and $P(\Omega) = 1$, then the probabilities of random events yield indeed the desired information about their degree of uncertainty: The closer $P(A)$ is to 0, the more unlikely is the occurrence of $A$, and the closer $P(A)$ is to 1, the more likely becomes the occurrence of $A$.

To formalize this intuitive approach, let for now $P = P(A)$ be a function on $\boldsymbol{E}$ with properties

I) $P(\emptyset) = 0$, $P(\Omega) = 1$, II) If $A \subseteq B$, then $P(A) \leq P(B)$.

As a corollary from these two properties we get the following property of $P$ :

III) For any event $A$, $0 \leq P(A) \leq 1$.

### 1.3.1 Classical Definition of Probability

The classical concept of probability is based on the following two assumptions:

1) The space $\boldsymbol{\Omega}$ of the elementary events is finite.

2) As a result of the underlying random experiment, each elementary event has the same probability to occur.

A random experiment with properties 1 and 2 is called a *Laplace random experiment.* Let $n$ be the total number of elementary events (i.e. the cardinality of $\boldsymbol{\Omega}$). Then any random event $A \subseteq \boldsymbol{\Omega}$ consisting of $m$ elementary events has probability

$$P(A) = m/n. \tag{1.4}$$

Let $\boldsymbol{\Omega} = \{a_1, a_2, \cdots, a_n\}$. Then every elementary event has probability

$$P(a_i) = 1/n, \quad i = 1, 2, ..., n.$$

Obviously, this definition of probability satisfies the properties I, II, and III listed above. The integer $m$ is said to be the number of *favorable cases* (for the occurrence of $A$), and $n$ is the number of *possible cases.*

The classical definition of probability arose in the Middle Ages to be able to determine the chances to win in various games of chance. Then formula (1.4) is applicable given that the players are honest and do not use marked cards or manipulated dice. For instance, what is the probability of the event $A$ that throwing a die yields an even number? In this case, $A = \{2, 4, 6\}$ so that $m = 3$ and $P(A) = 3/6 = 0.5$.

**Example 1.4** When throwing 3 dice, what is more likely, to achieve the total sum 9 (event $A_9$) or the total sum 10 (event $A_{10}$)? The corresponding sample space is

$$\boldsymbol{\Omega} = \{(i,j,k),\ 1 \leq i,j,k \leq 6\} \text{ with } n = 6^3 = 216$$

possible outcomes. The integers 9 and 10 can be represented a as sum of 3 positive integers in the following ways:

$$9 = 3+3+3 = 4+3+2 = 4+4+1 = 5+2+2 = 5+3+1 = 6+2+1,$$
$$10 = 4+3+3 = 4+4+2 = 5+3+2 = 5+4+1 = 6+2+2 = 6+3+1.$$

The sum 3+3+3 corresponds to the event $A_{333}$ = 'every die shows a 3' = $\{(3,3,3)\}$. The sum 4+3+2 corresponds to the event $A_{432}$ that one die shows a 4, another die a 3, and the remaining one a 2:

$$A_{432} = \{(2,3,4),(2,4,3),(3,2,4),(3,4,2),(4,2,3),(4,3,2)\}.$$

Analogously,

$$A_{441} = \{(1,4,4),(4,1,4),(4,4,1)\},\ \ A_{522} = \{(2,2,5),(2,5,2),(5,5,2),$$
$$A_{531} = \{(1,3,5),(1,5,3),(3,1,5),(3,5,1),(5,1,3),(5,3,1)\},$$
$$A_{621} = \{(1,2,6),(1,6,2),(2,1,6),(2,6,1),(6,1,2),(6,2,1)\}.$$

Corresponding to the given sum representations for 9 and 10, the numbers of favorable elementary events belonging to the events $A_9$ and $A_{10}$, respectively, are

$$m_A = 1+6+3+3+6+6 = 25, \quad m_B = 2+3+6+6+3+6 = 27.$$

Hence, the desired probabilities are:

$$P(A_9) = 25/216 = 0.116, \quad P(A_{10}) = 27/216 = 0.125.$$

The dice gamblers of the Middle Ages could not mathematically prove this result, but from their experience they knew that $P(A_9) < P(A_{10})$. □

**Example 1.5** $d$ dice are thrown at the same time.

What is the smallest number $d = d^*$ with property that the probability of the event $A =$ 'no die shows a 6' does not exceed 0.1?

The problem makes sense, since with increasing $d$ the probability $P(A)$ tends to 0, and if $d = 1$, then $P(A) = 5/6$. For $d \geq 1$, the corresponding space of elementary events $\mathbf{\Omega}$ has $n = 6^d$ elements, namely the vectors $(i_1, i_2, \cdots, i_d)$, where the $i_k$ are integers between 1 and 6. Amongst the $6^d$ elementary events those are favorable for the occurrence of $A$, where the $i_k$ only assume integers between 1 and 5. Hence, for the occurrence of $A$ exactly $5^d$ elementary events are favorable:

$$P(A) = 5^d/6^d.$$

The inequality $5^d/6^d \leq 0.1$ is equivalent to

$$d(\ln 5/6) \leq \ln(0.1) \ \text{ or } \ d(-0.1823) \leq -2.3026 \ \text{ or } \ d \geq \frac{2.3026}{0.1823} = 12.63.$$

Hence, $d^* = 13$. □

**Binomial Coefficient and Faculty** For solving the next examples, we need a result from elementary combinatorics: The number of possibilities to select subsets of $k$ different elements from a set of $n$ different elements, $1 \leq k \leq n$, is given by the *binomial coefficient* $\binom{n}{k}$, which is defined as

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!}, \quad 1 \leq k \leq n, \tag{1.5}$$

where $k!$ is the *faculty* of $k$: $k! = 1 \cdot 2 \cdots k$. By agreement

$$\binom{n}{0} = 1 \ \text{ and } \ 0! = 1.$$

The faculty of a positive integer has its own significance in combinatorics:

*There are $k!$ different possibilities to order a set of $k$ different objects.*

**Example 1.6** An optimist buys one ticket in a '6 out of 49' lottery and hopes for hitting the jackpot. What are his chances? There are

$$\binom{49}{6} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6!} = 13\,983\,816$$

different possibilities to select 6 numbers out of 49. Thus, one has to fill in almost 14 million tickets to make absolutely sure that the winning one is amongst them. It is $m = 1$ and $n = 13\,983\,816$. Hence, the probability $p_6$ of having picked the six 'correct' numbers is

$$p_6 = \frac{1}{13\,983\,816} = 0.0000000715. \qquad \square$$

The classical definition of probability satisfies the properties $P(\emptyset) = 0$ and $P(\mathbf{\Omega}) = 1$, since the impossible event $\emptyset$ does not contain any elementary events ($m = 0$) and the certain event $\mathbf{\Omega}$ comprises all elementary events ($m = n$).

Now, let $A$ and $B$ be two events containing $m_A$ and $m_B$ elementary events, respectively. If $A \subseteq B$, then $m_A \le m_B$ so that $P(A) \le P(B)$. If the events $A$ and $B$ are disjoint, then they have no elementary events in common so that $A \cup B$ contains $m_A + m_B$ elementary events. Hence

$$P(A \cup B) = \frac{m_A + m_B}{n} = \frac{m_A}{n} + \frac{m_B}{n} = P(A) + P(B)$$

or

$$P(A \cup B) = P(A) + P(B) \quad \text{if} \quad A \cap B = \emptyset. \tag{1.6}$$

More generally, if $A_1, A_2, \cdots, A_r$ are pairwise disjoint events, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_r) = P(A_1) + P(A_2) + \cdots + P(A_r), \quad A_i \cap A_k = \emptyset, \; i \ne k. \tag{1.7}$$

**Example 1.7** When participating in the lottery '6 out of 49' with one ticket, what is the probability of the event $A$ to have at least 4 correct numbers?

Let $A_i$ be the event of having got $i$ numbers correct. Then,

$$A = A_4 \cup A_5 \cup A_6.$$

$A_4, A_5$, and $A_6$ are pairwise disjoint events. (It is impossible that there are on one and the same ticket, say, exactly 4 and exactly 5 correct numbers.) Hence,

$$P(A) = P(A_4) + P(A_5) + P(A_6).$$

There are $\binom{6}{4} = 15$ possibilities to choose 4 numbers from the 6 'correct' ones. To each of these 15 choices there are

$$\binom{49-6}{6-4} = \binom{43}{2} = 903$$

possibilities to pick 2 numbers from the 43 'wrong' numbers. Therefore, favorable for the occurrence of $A_4$ are $m_4 = 15 \cdot 903 = 13\,545$ elementary events. Hence,

$$p_4 = P(A_4) = 13\,545/13\,983\,616 = 0.0009686336.$$

Analogously,

$$p_5 = P(A_5) = \frac{\binom{6}{5}\binom{49-6}{6-5}}{\binom{49}{6}} = \frac{6 \cdot 43}{\binom{49}{6}} = 0.0000184499.$$

Together with the result of example 1.6, $P(A) = p_4 + p_5 + p_6 = 0.0009871552$, i.e., almost 10 000 tickets have to be bought to achieve the desired result. □

### 1.3.2 Geometric Definition of Probability

The geometric definition of probability is subject to random experiments, in which every outcome has the same chance to occur (as with Laplace experiments), but the sample space $\mathbf{\Omega}$ is a bounded subset of the one, two or three dimensional Euklidian space (real line, plain, space). Hence, in each case $\mathbf{\Omega}$ is a nondenumerable set. In most applications, $\mathbf{\Omega}$ is a finite interval, a rectangular, a circle, a cube or a sphere.

Let $A \subseteq \mathbf{\Omega}$ be a random event. Then we denote by $\mu(A)$ the *measure* of $A$. For instance, if $\mathbf{\Omega}$ is a finite interval, then $\mu(\mathbf{\Omega})$ is the length of this interval. If $A$ is the union of disjoint subintervals of $\mathbf{\Omega}$, then $\mu(A)$ is the total length of these subintervals. (We do not consider subsets like the set of all irrational numbers in a finite interval.) If $\mathbf{\Omega}$ is a rectangular and $A$ is a circle embedded in this rectangular, then $\mu(A)$ is the area of this circle and so on. If $\mu$ is defined in this way, then

$$A \subseteq B \subseteq \mathbf{\Omega} \text{ implies } \mu(A) \le \mu(B) \le \mu(\mathbf{\Omega}).$$

Under the assumptions stated, a probability is assigned to every event $A \subseteq \mathbf{\Omega}$ by

$$P(A) = \frac{\mu(A)}{\mu(\mathbf{\Omega})}. \tag{1.8}$$

For disjoint events $A$ and $B$, $\mu(A \cup B) = \mu(A) + \mu(B)$ so that formulas (1.6) and (1.7) are true again. Analogously to the classical probability, $\mu(A)$ can be interpreted as the measure of all elementary events, which are favorable to the occurrence of $A$. With the given interpretation of the measure $\mu(\cdot)$, every elementary event, i.e. every point in $\mathbf{\Omega}$, has measure and probability 0 (different to the Laplace random experiment). (A point, whether at a line, in a plane or space has always extension 0 in all directions.) But the assumption "every elementary event has the same chance to occur" is not equivalent to the fact that every elementary event has probability 0. Rather, this assumption has to be understood in the following sense:

*All those random events, which have the same measure, have the same probability.*

Thus, never mind where the events (subsets of $\mathbf{\Omega}$) with the same measure are located in $\mathbf{\Omega}$ and however small their measure is, the outcome of the random experiment will be in any of these events with the same probability, i.e., no area in $\mathbf{\Omega}$ is preferred with regard to the occurrence of elementary events.

**Example 1.8** For the sake of a tensile test, a wire is clamped at its ends so that the free wire has a length of 400 *cm*. The wire is supposed to be homogeneous with regard to its physical parameters. Under these assumptions, the probability $p$ that the wire will tear up between 0 and 40 *cm* or 360 and 400 *cm* is

$$p = \frac{40+40}{400} = 0.2.$$

Repeated tensile tests will confirm or reject the assumption that the wire is indeed homogeneous. □

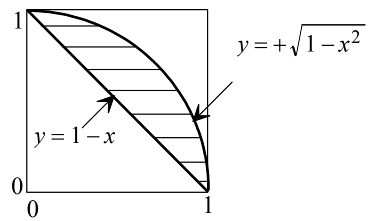



Figure 1.2 Illustration to example 1.9

**Example 1.9** Two numbers $x$ and $y$ are randomly picked from the interval $[0, 1]$. What is the probability that $x$ and $y$ satisfy both the conditions

$$x+y \geq 1 \text{ and } x^2+y^2 \leq 1?$$

Note: In this context, 'randomly' means that every number between 0 and 1 has the same chance of being picked.

In this case the sample space is the unit square $\mathbf{\Omega} = [0 \leq x \leq 1,\ 0 \leq y \leq 1]$, since an equivalent formulation of the problem is to pick at random a point out of the unit square, which is favorable for the occurrence of the event

$$A = \{(x,y);\ x+y \geq 1, x^2+y^2 \leq 1\}.$$

Figure 1.2 shows the area (hatched) given by $A$, whereas the 'possible area' $\mathbf{\Omega}$ is left white, but also includes the hatched area. Since $\mu(\mathbf{\Omega}) = 1$ and $\mu(A) = \pi/4 - 0.5$ (area of a quarter of a circle with radius 1 minus the area of the half of a unit square),

$$P(A) = \mu(A) \approx 0.2854.$$ □

**Example 1.10 (*Buffon's needle problem*)** At an even surface, parallel straight lines are drawn at a distance of $a$ *cm*. At this surface a needle of length $L$ is thrown, $L < a$. What is the probability of the event $A$ that the needle and a parallel intersect?

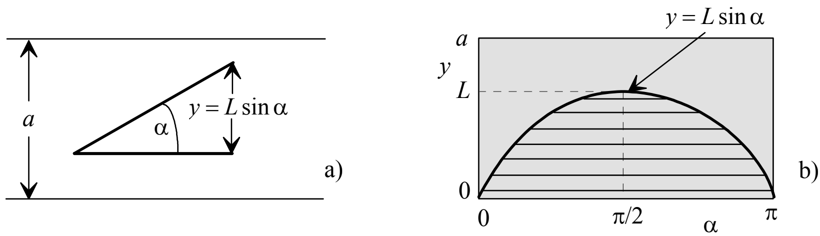



Figure 1.3 Illustration to example 1.10

The position of the needle at the surface is fully determined by its distance of its 'lower' endpoint to the 'upper' parallel and by its angle of inclination $\alpha$ to the parallels (Figure 1.3a), since a shift of the needle parallel to the lines obviously has no influence on the desired probability. Thus, the sample space is given by the rectangle

$$\mathbf{\Omega} = \{(y, \alpha),\ 0 \le y \le a,\ 0 \le \alpha \le \pi\}$$

with area $\mu(\mathbf{\Omega}) = a\,\pi$ (Figure 1.3b). Hence, Buffon's needle problem formally consists in randomly picking elementary events given by $(y, \alpha)$ from the rectangle $\mathbf{\Omega}$. Since the needle and the upper parallel intersect if and only if $y < L \sin\alpha$, the favorable area for the occurrence of $A$ is given by the hatched part in Figure 1.3b. The area of this part is

$$\mu(A) = \int_0^{\pi} L \sin\alpha\, d\alpha = L\,[-\cos\alpha]_0^{\pi} = L[1+1] = 2L.$$

Hence, the desired probability is $P(A) = 2L/a\pi$. □

### 1.3.3 Axiomatic Definition of Probability

The classical and the geometric concepts of probability are only applicable to very restricted classes of random experiments. But these concepts have illustrated which general properties a universally applicable probability definition should have:

**Definition 1.1** A function $P = P(A)$ on the set of all random events $\boldsymbol{E}$ with $\emptyset \in \boldsymbol{E}$ and $\mathbf{\Omega} \in \boldsymbol{E}$ is called *probability* if it has the following properties:

I) $P(\mathbf{\Omega}) = 1$.

II) For any $A \in \boldsymbol{E},\ 0 \le P(A) \le 1$.

III) For any sequence of disjoint events $A_1, A_2, \ldots,$ i.e., $A_i \cap A_j = \emptyset$ for $i \ne j$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{1.9}$$

●

Property III makes sense only if with $A_i \in \boldsymbol{E}$ the union $\bigcup_{i=1}^{\infty} A_i$ is also an element of $\boldsymbol{E}$. Hence we assume that the set of all random events $\boldsymbol{E}$ is a $\sigma$–algebra:

**Definition 1.2** Any set of random events $\boldsymbol{E}$ is called a $\sigma-$*algebra* if
1) $\boldsymbol{\Omega} \in \boldsymbol{E}$.
2) If $A \in \boldsymbol{E}$, then $\overline{A} \in \boldsymbol{E}$. In particular, $\overline{\boldsymbol{\Omega}} = \emptyset \in \boldsymbol{E}$.
3) For any sequence $A_1, A_2, ...$ with $A_i \in \boldsymbol{E}$, the union $\bigcup_{i=1}^{\infty} A_i$ is also a random event, i.e.,

$$\bigcup_{i=1}^{\infty} A_i \in \boldsymbol{E}.$$

$[\boldsymbol{\Omega}, \boldsymbol{E}]$ is called a *measurable space*, and $[\boldsymbol{\Omega}, \boldsymbol{E}, P]$ is called a *probability space*. ●

Note: In case of a finite or a countably infinite set $\boldsymbol{\Omega}$, the set $\boldsymbol{E}$ is usually the *power set* of $\boldsymbol{\Omega}$, i.e. the set of all subsets of $\boldsymbol{\Omega}$. A power set is, of course, always a $\sigma-$algebra. In this book, taking into account its applied orientation, specifying explicitly the underlying $\sigma-$ algebra is usually not necessary. $[\boldsymbol{\Omega}, \boldsymbol{E}]$ is called a *measurable space,* since to any random event $A \in \boldsymbol{E}$ a *measure*, namely its probability, can be assigned. In view of the de Morgan rules (1.1): If $A$ and $B$ are elements of $\boldsymbol{E}$, then $A \cap B$ as well.

Given that $\boldsymbol{E}$ is a $\sigma-$algebra, properties I–III of definition 1.1 imply all the properties of the probability functions, which we found useful in sections 1.3.1 and 1.3.2:
a) Let $A_i = \emptyset$ for $i = n+1, n+2, \cdots$. Then, from III),

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i), \quad A_i \cap A_j = \emptyset, \ i \neq j, \ i,j = 1, 2, \cdots, n. \tag{1.10}$$

In particular, letting $n = 2$ and $A = A_1$, $B = A_2$, this formula implies

$$P(A \cup B) = P(A) + P(B) \ \text{ if } \ A \cap B = \emptyset. \tag{1.11}$$

With $B = \overline{A}$, taking into account $\boldsymbol{\Omega} = A \cup \overline{A}$ and $P(\boldsymbol{\Omega}) = 1$, formula (1.11) yields

$$P(A) + P(\overline{A}) = 1 \ \text{ or } \ P(\overline{A}) = 1 - P(A). \tag{1.12}$$

Applying (1.12) with $A = \boldsymbol{\Omega}$ yields $P(\boldsymbol{\Omega}) + P(\emptyset) = 1$, so that

$$P(\emptyset) = 0, \quad P(\boldsymbol{\Omega}) = 1. \tag{1.13}$$

Note that $P(\boldsymbol{\Omega}) = 1$ is part of definition 1.1.

b) If $A$ and $B$ are two events with $A \subseteq B$, then $B$ can be represented as $B = A \cup (B \backslash A)$. Since $A$ and $B \backslash A$ are disjoint, by (1.11), $P(B) = P(A) + P(B \backslash A)$ or, equivalently,

$$P(B \backslash A) = P(B) - P(A) \ \text{ if } \ A \subseteq B. \tag{1.14}$$

Therefore,

$$P(A) \leq P(B) \ \text{ if } \ A \subseteq B. \tag{1.15}$$

c) For any events $A$ and $B$, the event $A \cup B$ can be represented as follows (Figure 1.1)

$$A \cup B = \{A \backslash A \cap B)\} \cup \{B \backslash (A \cap B)\} \cup (A \cap B).$$

In this representation, the three events combined by '$\cup$' are disjoint. Hence, by (1.10) with $n = 3$:

$$PA \cup B) = P(\{A \backslash A \cap B)\}) + P(\{B \backslash (A \cap B)\}) + P(A \cap B).$$

On the other hand, since $(A \cap B) \subseteq A$ and $(A \cap B) \subseteq B$, from (1.14),

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1.16}$$

Given any 3 events *A, B*, and *C,* the probability of the event $A \cup B \cup C$ can be determined by replacing in (1.16) $A$ with $A \cup B$ and $B$ with $C$. This yields

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ +P(A \cap B \cap C) \tag{1.17}$$

d) For any $n$ events $A_1, A_2, ..., A_n$ one obtains by repeated application of (1.16) (more exactly, by induction) the *Inclusion-Exclusion Formula* or the *Formula of Poincaré* for the probability of the event $A_1 \cup A_2 \cup \cdots \cup A_n$ :

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} R_k \tag{1.18}$$

with
$$R_k = \sum_{(i_1 < i_2 < \cdots < i_k)}^{n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}),$$

where the summation runs over all $k$-dimensional vectors $(i_1, i_2, ..., i_k)$ out of the set $\{1, 2, ..., n\}$ with $1 \le i_1 < i_2 < \cdots < i_k \le n$ and $k = 1, 2, ..., n$. The sum representing $R_k$ has exactly $\binom{n}{k}$ terms, so that the total number of terms in (1.18) is

$$\sum_{k=1}^{n} \binom{n}{k} = 2^n - 1.$$

For instance, if $n = 3$, then the $R_k$ in (1.18) are

$$R_1 = P(A_1) + P(A_2) + P(A_3),$$
$$R_2 = P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3),$$
$$R_3 = P(A_1 \cap A_2 \cap A_3).$$

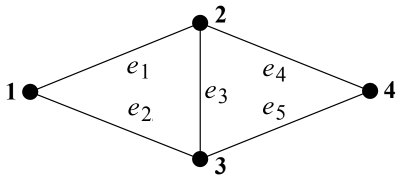



Figure 1.4 Computer network with 4 computers

**Example 1.11** Figure 1.4 shows a simple local computer network. Computers are located at nodes 1, 2, 3, and 4. The transmission of data between the computers is possible via cables $e_1, e_2, \cdots, e_5$, which link the four computers. Cable $e_i$ is available, i.e. in a position to transfer information, with probability $p_i$ and unavailable (e.g. under maintenance, waiting for maintenance, waiting for replacement for having been stolen) with probability $q_i = 1 - p_i$, $i = 1, 2, ..., 5$.

What is the probability of the event $A$ that the computer at node 1 can transfer data to the computer at node 4 via one or more paths (chains) of available edges which connect node 1 to node 4? There are four potential candidates for such paths:

$$w_1 = \{e_1, e_4\},\ w_2 = \{e_2, e_5\},\ w_3 = \{e_1, e_3, e_5\},\ w_4 = \{e_2, e_3, e_4\}.$$

Let $A_i$ be the event that all edges in path $w_i$ are available, $i = 1, 2, 3, 4.$ Then event $A$ occurs when at least one of these four events occurs. Hence, $A$ can be represented as

$$A = A_1 \cup A_2 \cup A_3 \cup A_4.$$

The $A_i$ are not disjoint. Hence we apply the inclusion-exclusion formula (1.11) for representing $A$:

$$P(A) = P(A_1 \cup A_2 \cup A_3 \cup A_4) = R_1 - R_2 + R_3 - R_4$$

with

$$R_1 = P(A_1) + P(A_2) + P(A_3) + P(A_4),$$

$$R_2 = P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_1 \cap A_4) + P(A_2 \cap A_3) + P(A_2 \cap A_4)$$
$$+ P(A_2 \cap A_4) + P(A_3 \cap A_4),$$

$$R_3 = P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_4) + P(A_1 \cap A_3 \cap A_4) + P(A_2 \cap A_3 \cap A_4),$$

$$R_4 = P(A_1 \cap A_2 \cap A_3 \cap A_4).$$

The event $A_1 \cap A_2$ means that both the cables in $A_1$ and in $A_2$ are operating. Thus, to the event $A_1 \cap A_2$ there belongs the set of cables $\{e_1, e_2, e_4, e_5\}$. Hence, the notation $P(A_1 \cap A_2) = p_{1245}$ will be used. To the event $A_1 \cap A_2 \cap A_3$ there belongs the set of cables $\{e_1, e_2, e_3, e_4, e_5\}$: $P(A_1 \cap A_2 \cap A_3) = p_{12345}$. If this way of notation is applied to all other probabilities occurring in the $R_i$, then

$$R_1 = p_{14} + p_{25} + p_{135} + p_{234},$$

$$R_2 = p_{1245} + p_{1345} + p_{1234} + p_{1235} + p_{2345} + p_{12345},$$

$$R_3 = p_{12345} + p_{12345} + p_{12345} + p_{12345},\quad R_4 = p_{12345}.$$

The desired probability is

$$P(A) = p_{14} + p_{25} + p_{135} + p_{234} - p_{1245} - p_{1345} - p_{1234} - p_{1235} - p_{2345} + 3p_{12345}.$$

In section 1.4.2, an additional assumption on the operation modus of the cables will be imposed which enables the calculation of $P(A)$ only on the basis of the $p_i$. □

### 1.3.4 Relative Frequency

The probabilities of random events are usually unknown. However, they can be estimated by their relative frequencies. If in a series of $n$ repetitions of one and the same random experiment the event $A$ has been observed exactly $m = m(A)$ times, then the *relative frequency* of $A$ is given by

$$\hat{p}_n(A) = \frac{m(A)}{n}. \tag{1.19}$$

Generally, the relative frequency of $A$ tends to $P(A)$ as $n$ increases:

$$\lim_{n\to\infty} \hat{p}_n(A) = P(A). \tag{1.20}$$

Thus, the probability of $A$ can be estimated with any required level of accuracy from its relative frequency by sufficiently frequently repeating the random experiment (for the theoretical background see section 5.2.2). Empirical verifications of the limit relation (1.20) were aleady given in the introduction by the coin experiments of *Buffon* and *Pearson*. Without the validity of (1.20) the gamblers in the Middle Ages would not have been in a position to empirically verify that, when throwing three dice, the chance to obtain sum 9 is lower than the chance to obtain sum 10 (example 1.4).

It is interesting that the relationship (1.20) in connection with Buffon's needle problem (example 1.10) allows to estimate the number $\pi$ with any desired degree of accuracy. To do this, in the formula $P(A) = 2L/\pi a$ the probability $P(A)$ is replaced with the relative frequency $\hat{p}_n(A)$ for the occurrence of $A$ in a series of $n$ needle throwings. This gives for $\pi$ the estimate

$$\hat{\pi}_n = \frac{2L}{a\,\hat{p}_n(A)}.$$

*Lazzarini* (1901) threw the needle $n = 3408$ times and got for $\pi$ the estimate

$$\hat{\pi}_{3408} = 3.141529,$$

i.e., the first six figures are the exact ones. The approximate calculation of $\pi$ was one of the first examples how to solve deterministic problems by probabilistic methods. Nowadays, nobody needs to throw a needle manually several tousand times. Computers 'simulate' random experiments of this simple structure many thousand times in a twinkling of an eye.

The reader may object that the approximate calculation of probabilities of all random events by their relative frequency is practically not possible, in particular, if the sample space is not finite. However, depending on the respective random experiment, the probabilities of all its elementary events are frequently given by a unifying mathematical pattern (model). For instance, the probability that the random number of traffic accidents occurring in a specific area during a year is equal to $k$ can frequently be determined by the formula

$$p_k = \frac{\lambda^k}{k!}e^{-\lambda}; \;\; k = 0, 1, \dots,$$

where $\lambda$ is the average number of traffic accidents which occur a year in that area. Hence, for determining all infinitely many probabilities $p_0, p_1, \dots$, only the parameter $\lambda$ has to be estimated. This is done by counting the number $x_i$ of traffic accidents occurring in year $i$ over a period of $n$ years and determining the arithmetic mean

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Defining and discussing mathematical models for the calculation of the probabilities of random events is the subject of chapter 2.

## 1.4 CONDITIONAL PROBABILITY AND INDEPENDENCE OF RANDOM EVENTS

### 1.4.1 Conditional Probability

Two random events $A$ and $B$ can depend on each other in the following sense: The occurrence of $B$ will change the probability of the occurrence of $A$ and vice versa. Hence, the additional piece of information '$B$ has occurred' should be used in order to predict the probability of the occurrence of $A$ more precisely. If one has to determine the probability that a device does not fail during its guarantee period (event $A$), then this probability may depend on the manufacturer of the device (event $B$) if there are several of them who produce the same type. The probability of having a sunny day on 21 August (event $A$) will increase if there is a sunny day on 20 August (event $B$) in view of the inertia of weather patterns. The probability of attracting a certain disease (event $A$) will usually be larger than average if there was/is a family member, who had suffered from this disease (event $B$). If $A$ is the random event to spot a leopard in a certain area of a National Park during a safari, then the probability of $A$ increases if it is known that there are baboons in this area (event $B$).

Let us now consider some numerical examples to illustrate how to define the probability of the occurrence of an event $A$ given that another event $B$ has occurred.

**Example 1.12** A gambler throws the dice 1 and 2 simultaneously. What is the probability that die 1 shows a 6 (event $A$) on condition that both dice showed an even number (event $B$). This probability will be denoted as $P(A|B)$. The sample space is

$$\mathbf{\Omega} = \{(i,j);\ i,j = 1, 2, ..., 6\}.$$

In terms of the elementary events $(i,j)$, the events $A$ and $B$ are given by

$$A = \{(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\},$$

$$B = \{(2,2), (2,4), (2,6), (4,2), (4,4), (4,6), (6,2), (6,4), (6,6)\}.$$

Hence,

$$P(A) = 6/36 \quad \text{and} \quad P(B) = 9/36.$$

On condition '$B$ has occurred' the sample space $\mathbf{\Omega}$ reduces to the 9 elementary events given by $B$. From these 9, only the 3 elementary events in the conjunction

$$A \cap B = \{(6,2), (6,4), (6,6)\}$$

are favorable for the occurrence of $A$: Therefore,

$$P(A|B) = 3/9.$$

The following representation shows the general structure of $P(A|B)$:

$$P(A|B) = 1/3 = \frac{3/36}{9/36} = \frac{P(A \cap B)}{P(B)}. \qquad \square$$

**Example 1.13** In a bowl there are two white and two red marbles. The numbers 1 and 2 are assigned to the white marbles and the numbers 3 and 4 are assigned to the red marbles. Two marbles are one after the other randomly picked from the bowl. Find the probability of the event $A$ that one of the drawn marbles is white and the other red given the event $B$ that the first drawn marble is white.

The sample space consists of $4 \cdot 3 = 12$ elementary events:

$$\Omega = \{(i,j);\ i \neq j,\ i,j = 1,2,3,4\}.$$

The events $A$ and $B$ are given by the following sets of elementary events:

$$A = \{(1,3), (1,4), (2,3), (2,4), (3,1), (3,2), (4,1), (4,2)\},$$

$$B = \{(1,2), (1,3), (1,4), (2,1), (2,3), (2,4)\}.$$

Hence,

$$P(A) = 8/12 = 2/3 \quad \text{and} \quad P(B) = 6/12 = 1/2.$$

Since it is known that event $B$ has happened, the space of possible elementary events is given by $B$. Hence, the elementary events which are favorable for the occurrence of event $A$ are given by the conjunction

$$A \cap B = \{(1,3), (1,4), (2,3), (2,4)\}.$$

This yields

$$P(A|B) = \frac{4}{6} = \frac{2}{3} = \frac{4/12}{6/12} = \frac{P(A \cap B)}{P(B)}.$$

For the sake of arriving at the general structure of $P(A|B)$, solution of the problem had been unnecessarily complicated. The problem is namely quickly solved as follows: If the first drawn marble is white (event $B$), then there are one white and two red marbles left in the bowl. Event $A$ occurs if one of the red marbles will be drawn, i.e., $P(A|B) = 2/3$. □

**Example 1.14** The lifetimes of $n = 1000$ electronic units had been tested. 205 units failed in the interval $[0, 500\,h)$, 180 units failed in the interval $[500, 600\,h)$, and the remaining 615 units failed after $600\,h$. Let $A$ be the event that a unit fails in the interval $[500, 600\,h)$, and $B$ be the event that a unit fails after a lifetime of at least $500\,h$. By formula (1.19) with $n = 1000$, the relative frequencies for the occurrence of events $A$ and $B$ are

$$\hat{p}_n(A) = \frac{m(A)}{n} = \frac{180}{1000}, \quad \hat{p}_n(B) = \frac{m(B)}{n} = \frac{1000 - 205}{1000} = 0.795.$$

What is the relative frequency $\hat{p}_n(A|B)$ of the event $A$ on condition that event $B$ has occurred?

Under this condition, only the 795 units, which have survived the first $500\,h$, need to be taken into account. From these 795 units, 180 fail in $[500, 600\,h)$. Therefore,

$$\hat{p}_n(A|B) = \frac{180}{795} = 0.2264.$$

Since $A \subseteq B$, i.e. the occurrence of $A$ implies the occurrence of $B$, event $A$ satisfies $A = A \cap B$. Hence, the 'conditional relative frequency' $\hat{p}_n(A|B)$ can be written as

$$\hat{p}_n(A|B) = \frac{m(A \cap B)}{m(B)} = \frac{\frac{m(A\cap B)}{n}}{\frac{m(B)}{n}}. \tag{1.21}$$

By (1.20), the relative frequencies $\frac{m(A\cap B)}{n}$ and $\frac{m(B)}{n}$ tend to $P(A \cap B)$ and $P(B)$ as $n \to \infty$, respectively. Thus, the conditional probability of $A$ given $B$ has again the structure we know from the previous examples:

$$\lim_{n\to\infty} \hat{p}_n(A|B) = P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

□

Now it is no longer surprising that the probability of '$A$ on condition $B$' or, equivalently, the probability of '$A$ given $B$' is defined as follows.

**Definition 1.3** Let $A$ and $B$ be two events with $P(B) > 0$. Then the *probability* of *A on condition B* is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1.22}$$

●

Note: $P(A|B)$ is also denoted as the *probability of A given B,* the *conditional probability of A on condition B,* or *the conditional probability of A given B*. Of course, in (1.22) the roles of $A$ and $B$ can be changed.

If $A$ and $B$ are arbitrary random events, formula (1.22) implies a *product formula* for the probability $P(A \cap B)$ of the joint occurrence of arbitrary events $A$ and $B$:

$$P(A \cap B) = P(A|B)\ P(B) \quad \text{or} \quad P(A \cap B) = P(B|A)\ P(A). \tag{1.23}$$

**Example 1.15** In a bowl there are three white and two red marbles. Two marbles are randomly taken out one after the other. What is the probability that both of these marbles are red?

Let be $A$ and $B$ be the events that the first and the second, respectively, of the chosen marbles are red. Hence, the probability $P(A \cap B)$ has to be determined. The probability of $A$ is equal to $P(A) = 2/5$. On condition $A$, there are 3 white and 1 red marble in the bowl. Hence, $P(B|A) = 1/4$ so that

$$P(A \cap B) = P(B|A)\,P(A) = \tfrac{1}{4} \cdot \tfrac{2}{5} = 0.1.$$

□

**Example 1.16** In a study, data from a sample of 12 000 persons had been collected. 4800 persons in this sample were obese and 3600 suffered from diabetes 2. From the diabetes sufferers, 2700 were obese. A person is randomly selected from the sample of 12 000 persons. It happens to be Max. Let $A$ be the event that Max is obese, and $B$ be the event that Max has diabetes 2. Then

$$P(A) = 0.4,\ P(B) = 0.3, \text{ and } P(A|B) = 2700/3600 = 0.75.$$

Hence, the probability that Max is both obese and a diabetes 2 sufferer is, by (1.22),

$$P(A \cap B) = P(A|B)P(B) = 0.75 \cdot 0.3 = 0.225.$$

2) To see whether being obese increases the probability of contracting diabetes 2, the probability $P(B|A)$ has to be determined: From the right equation of (1.23),

$$P(A \cap B) = 0.225 = P(B|A)P(A) = P(B|A) \cdot 0.4.$$

Hence, $P(B|A) = 0.5625$. Thus, based on this study, being obese increases the probability of contracting diabetes 2. □

### 1.4.2 Total Probability Rule and Bayes' Theorem

Frequently several mutually exclusive conditions have influence on the occurrence of a random event $A$. The whole of these conditions are known, but it is not known, which of these conditions is taking effect. However, the probabilities are known which of these conditions affects the occurrence of $A$ at the time point of interest. Under these assumptions, a formula for the occurrence of $A$ will be derived. But next the procedure is illustrated by an example.

**Example 1.17** A machine is subject to two stress levels 1 (event $B_1$) and 2 (event $B_2$) with respective probabilities 0.8 and 0.2. Stress levels can be determined e.g. by different production conditions as speed, pressu,re or humidity. It is supposed that the stress level does not change during a fixed working period (hour, day). Given stress level 1 or 2, the machine will fail during a working period with probability 0.3 or 0.6, respectively. Hence,

$$P(A|B_1) = 0.3,\ P(A|B_2) = 0.6.$$

Since the events $B_1$ and $B_2$ are disjoint (mutually exclusive) and $\mathbf{\Omega} = B_1 \cup B_2$ is the certain event, $A$ can be represented as

$$A = A \cap \mathbf{\Omega} = A \cap (B_1 \cup B_2) = (A \cap B_1) \cup (A \cap B_2).$$

The events $A \cap B_1$ and $A \cap B_2$ are disjoint so that by formula (1.11)

$$P(A) = P(A \cap B_1) + P(A \cap B_2).$$

By applying (1.23) to each of the two terms on the right-hand side of this formula,

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) \\ &= 0.3 \cdot 0.8 + 0.6 \cdot 0.2 = 0.36. \end{aligned}$$

Thus, without information on the respective stress level, the failure probability of the machine in the working period is 0.36. □

Now the principle, illustrated by this example, is formulated more generally:

**Definition 1.4** The set of random events $\{B_1, B_2, ..., B_n,\ n \le \infty\}$ is an *exhaustive set of random events* for $\mathbf{\Omega}$ if

$$\mathbf{\Omega} = \bigcup_{i=1}^{n} B_i\,,$$

and it is a mutually disjoint set of events if

$$B_i \cap B_j = \emptyset,\ \ i \ne j,\ \ i,j = 1, 2, ..., n.$$

A mutually disjoint and exhaustive (for $\mathbf{\Omega}$) set of events is called a *partition* of $\mathbf{\Omega}$. ●

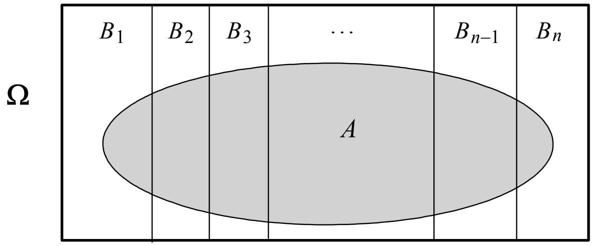



Figure 1.5 Partition of a sample space

Let $\{B_1, B_2, ..., B_n\}$ be an exhaustive and mutually disjoint set of events with property $P(B_i) > 0$ for all $i = 1, 2, ..., n$, and let $A$ be an event with $P(A) > 0$. Then $A$ can be represented as follows (see Figure 1.5):

$$P(A) = \bigcup_{i=1}^{n} (A \cap B_i).$$

Since the $B_i$ are disjoint, the conjunctions $A \cap B_i$ are disjoint as well. Formula (1.10) is applicable and yields $P(A) = \Sigma_{i=1}^{n} P(A \cap B_i)$. Now formula (1.23) applied to all $n$ probabilities $P(A \cap B_i)$ yields

$$P(A) = \Sigma_{i=1}^{n} P(A|B_i)\,P(B_i). \tag{1.24}$$

This result is called the *Formula of total probability* or the *Total probability rule.*

Moreover, formulas (1.22) and (1.23) yield

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

If $P(A)$ is replaced with its representation (1.24), then

$$P(B_i|A) = \frac{P(A|B_i)\,P(B_i)}{\Sigma_{i=1}^{n} P(A|B_i)\,P(B_i)},\quad i = 1, 2, ..., n. \tag{1.25}$$

Formula (1.25) is called *Bayes' theorem* or *Formula of Bayes*. For obvious reasons, the probabilities $P(B_i)$ are called *a priori probabilities* and the conditional probabilities $P(B_i\,|\,A)$ *a posteriori probabilities.*

**Example 1.18** The manufacturers $M_1, M_2$, and $M_3$ delivered to a supermarket a total of 1000 fluorescent tubes of the same type with shares 200, 300, and 500, respectively. In these shares, there are in this order 12, 9, and 5 defective tubes.

1) What is the probability that a randomly chosen tube is not defective?

2) What is the probability that a defective tube had been produced by $M_i$, $i = 1, 2, 3$?

Let events $A$ and $B_i$ be introduced as follows:

$A$ = 'A tube, randomly chosen from the whole delivery, is not defective.'

$B_i$ = 'A tube, randomly chosen from the whole delivery, is from $M_i$, $i = 1, 2, 3$.'

According to the figures given:

$$P(B_1) = 0.2,\ P(B_2) = 0.3,\ P(B_3) = 0.5,$$

$$P(A|B_1) = 12/200 = 0.06,\ P(A|B_2) = 9/300 = 0.03,\ P(A|B_3) = 5/500 = 0.01.$$

$\{B_1, B_2, B_3\}$ is a set of exhaustive and mutually disjoint events, since there are no other manufacturers delivering tubes of this brand to that supermarket and no two manufacturers can have produced one and the same tube.

1) Formula (1.23) yields

$$P(A) = 0.06 \cdot 0.2 + 0.03 \cdot 0.3 + 0.01 \cdot 0.5 = 0.026.$$

2) Bayes' theorem (1.25) gives the desired probabilities:

$$P(B_1|A) = \frac{P(A|B_1)\,P(B_1)}{P(A)} = \frac{0.06 \cdot 0.2}{0.026} = 0.4615,$$

$$P(B_2|A) = \frac{P(A|B_2)\,P(B_2)}{P(A)} = \frac{0.03 \cdot 0.3}{0.026} = 0.3462,$$

$$P(B_3|A) = \frac{P(A|B_3)\,P(B_3)}{P(A)} = \frac{0.01 \cdot 0.5}{0.026} = 0.1923.$$

Thus, despite having by far the largest proportion of tubes in the delivery, the high quality of tubes from manufacturer $M_3$ guarantees that a defective tube is most likely not produced by this manufacturer. □

**Example 1.19** 1% of the population in a country are HIV-positive. A test procedure for diagnosing whether a person is HIV-positive indicates with probability 0.98 that the person is HIV-positive if indeed he/she is HIV-positive, and with probability 0.96 that this person is not HIV-positve if he/she is not HIV-positive.

1) What is the probability that a test person is HIV-positive if the test indicates that?

To solve the problem, random events $A$ and $B$ are introduced:

$A$ = 'The test indicates that a person is HIV-positive.'

$B$ = 'A test person is HIV-positive.'

Then, from the figures given,

$$P(B) = 0.01,\ \ P(\overline{B}) = 0.99$$

$$P(A|B) = 0.98,\ \ P(\overline{A}|B) = 0.02,\ \ P(\overline{A}|\overline{B}) = 0.96,\ \ P(A|\overline{B}) = 0.04.$$

Since $\{B, \overline{B}\}$ is an exhaustive and disjoint set of events, the total probability rule (1.23) is applicable to determining $P(A)$:

$$P(A) = P(A|B)\,P(B) + P(A|\overline{B})\ P(\overline{B}) = 0.98 \cdot 0.01 + 0.04 \cdot 0.99 = 0.0494.$$

Bayes' theorem (1.24) yields the desired probability $P(B|A)$:

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} = \frac{0.98 \cdot 0.01}{0.0494} = 0.1984.$$

Although the initial parameters of the test look acceptable, this result is quite unsatisfactory: In view of $P(\overline{B}|A) = 0.8016$, about 80% HIV-negative test persons will be shocked to learn that the test procedure indicates they are HIV-positive. In such a situation the test has to be repeated several times. The reason for this unsatisfactory numerical result is that only a small percentage of the population is HIV-positive.

2) The probability that a person is HIV-negative if the test procedure indicates this is

$$P(\overline{B}|\overline{A}) = \frac{P(\overline{A}|\overline{B})\,P(\overline{B})}{P(\overline{A})} = \frac{0.96 \cdot 0.99}{1 - 0.0494} = 0.99979\,.$$

This result is, of course, an excellent feature of the test. □

### 1.4.3 Independent Random Events

If a die is thrown twice, then the result of the first throw does not influence the result of the second throw and vice versa. If you have not won in the weekly lottery during the past 20 years, then this bad luck will not increase or decrease your chance to win in the lottery the following week. An aircraft crash over the Pacific for technical reasons has no connection to the crash of an aircraft over the Atlantic for technical reasons the same day. Thus, there are random events which do not at all influence each other. Events like that are called independent (of each other). Of course, for a quantitative probabilistic analysis a more accurate definition is required.

If the occurrence of a random event $B$ has no influence on the occurrence of a random event $A$, then the probability of the occurrence of $A$ will not be changed by the additional information that $B$ has occurred, i.e.

$$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1.26}$$

This motivates the definition of independent random events:

**Definition 1.5**: Two random events $A$ and $B$ are called *independent* if

$$P(A \cap B) = P(A)\,P(B)\,. \tag{1.27}$$

●

This is the *product formula* for independent events $A$ and $B$. Obviously, (1.27) is also valid for $P(B) = 0$ and/or $P(A) = 0$. Hence, defining independence of two random events by (1.27) is preferred to defining independence by formula (1.26).

If $A$ and $B$ are independent random events, then the pairs $A$ and $\overline{B}$, $\overline{A}$ and $B$, as well as $\overline{A}$ and $\overline{B}$ are independent, too. That means relation (1.27) implies, e.g.,

$$P(A \cap \overline{B}) = P(A)\,P(\overline{B}).$$

This can be proved as follows:

$$\begin{aligned} P(A \cap \overline{B}) &= P(A \cap (\Omega \backslash B)) = P((A \cap \Omega)\backslash(A \cap B)) = P(A\backslash(A \cap B)) \\ &= P(A) - P(A \cap B) = P(A) - P(A)P(B) \\ &= P(A)[1 - P(B)] = P(A)\,P(\overline{B}). \end{aligned}$$

The generalization of the independence property to more than two random events is not obvious. The pairwise independence between $n \geq 2$ events is defined as follows: The events $A_1, A_2, ..., A_n$ are called *pairwise independent* if for each pair $(A_i, A_j)$

$$P(A_i \cap A_j) = P(A_i)\,P(A_j), \quad i \neq j,\ i,j = 1,2,...,n.$$

A more general definition of the independence of $n$ events is the following one:

**Definition 1.6** The random events $A_1, A_2, ..., A_n$ are called *completely independent* or simply *independent* if for all $k = 2, 3, ..., n$,

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})\,P(A_{i_2}) \cdots P(A_{i_k}) \tag{1.28}$$

for any subset $\{A_{i_1}, A_{i_2}, ..., A_{i_k}\}$ of $\{A_1, A_2, ..., A_n\}$ with $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. ●

Thus, to verify the complete independence of $n$ random events, one has to check

$$\sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1$$

conditions. Luckily, in most applications it is sufficient to verify the case $k = n$:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)\,P(A_2) \cdots P(A_n). \tag{1.29}$$

The complete independence is a stronger property than the pairwise independence. For this reason it is interesting to consider an example, in which the $A_1, A_2, ..., A_n$ are pairwise independent, but not complete independent.

**Example 1.20** The dice $D_1$ and $D_2$ are thrown. The corresponding sample space consists of 36 elementary events: $\mathbf{\Omega} = \{(i,j);\ i,j = 1,2,...,6\}$. Let

$A_1$ = '$D_1$ shows a 1' = $\{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)\}$,

$A_2$ = '$D_2$ shows a 1' = $\{(1,1), (2,1), (3,1), (4,1), (5,1), (6,1)\}$,

$A_3$ = 'both $D_1$ and $D_2$ show the same number' = '$\{(i,i),\ i = 1,2,...,6)\}$.'

Since the $A_i$ each contain 6 elementary events,

$$P(A_1) = P(A_2) = P(A_3) = 1/6\,.$$

The $A_i$ have only one elementary event in common, namely $(1, 1)$. Hence,

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \tfrac{1}{6} \cdot \tfrac{1}{6} = \tfrac{1}{36}.$$

Therefore, the $A_i$ are pairwise independent. However, there is

$$A_1 \cap A_2 \cap A_3 = \{(1, 1)\}.$$

Hence,

$$P(A_1 \cap A_2 \cap A_3) = \tfrac{1}{36} \neq P(A_1)P(A_2)P(A_3) = \tfrac{1}{6} \cdot \tfrac{1}{6} \cdot \frac{1}{6} = \tfrac{1}{216}. \qquad \square$$

**Example 1.21** (*Chevalier de Méré*) What is more likely: 1) to get at least one 6, when throwing four dice simultaneously (event $A$), or 2) to get the outcome (6,6) at least once, when throwing two dice 24 times simultaneously (event $B$)?

The complementary events to $A$ and $B$ are:

$\overline{A}$ = 'none of the dice shows a 6, when four dice are thrown simultaneously,'

$\overline{B}$ = 'the outcome (6,6) does not occur, when two dice are thrown 24 times.'

1) Both the four results obtained by throwing four or two dice and the results by repeatedly throwing two dice are independent of each other. Hence, since the probability to get no 6, when throwing one die, is 5/6, formula (1.29) with $n = 4$ yields

$$P(\overline{A}) = (5/6)^4.$$

The probability, not to get the result (6,6) when throwing two dice, is 35/36. Hence, formula (1.29) yields with $n = 24$ the probability

$$P(\overline{B}) = (35/36)^{24}.$$

Thus, the desired probabilities are

$$P(A) = 1 - (5/6)^4 \approx 0.518, \quad P(B) = 1 - (35/36)^{24} \approx 0.491. \qquad \square$$

**Example 1.22** In a set of traffic lights, the color 'red' (as well as green and yellow) is indicated by two bulbs which operate independently of each other. Color 'red' is clearly visible if at least one bulb is operating.

What is the probability that in the time interval [0, 200 *hours*] color 'red' is visible if it is known that a bulb survives this interval with probability 0.95 ?

To answer this question, let

$A$ = 'bulb 1 does not fail in [0, 200],' $B$ = 'bulb 2 does not fail in [0, 200].'

The event of interest is

$$C = A \cup B = \text{'red light is clearly visible in [0, 200].'}$$

By formula (1.16),

$$P(C) = P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Since $A$ and $B$ are independent,

$$P(C) = P(A) + P(B) - P(A)\,P(B) = 0.95 + 0.95 - (0.95)^2.$$

Thus, the desired probability is

$$P(C) = 0.9975.$$

Another possibility of solving this problem is to apply the *Rules of de Morgan* (1.1):

$$P(\overline{C}) = P(\overline{A \cup B}) = P(\overline{A} \cap \overline{B}) = P(\overline{A})\,P(\overline{B}) = (1 - 0.95)(1 - 0.95)$$
$$= 0.0025$$

so that $P(C) = 1 - P(\overline{C}) = 0.9975.$ □

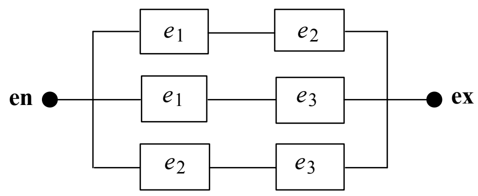



Figure 1.6 Diagram of a '2 out of 3-system'

**Example 1.23** ('2 out of 3 system') A system $S$ consists of 3 independently operating subsystems $S_1$, $S_2$, and $S_3$. The system operates if and only if at least 2 of its subsystems operate. Figure 1.6 illustrates the situation: $S$ operates if there is at least one path with two operating subsystems (symbolized by rectangles) from the entrance node **en** to the exit node **ex**. As an application may serve the following one: The pressure in a high-pressure tank is indicated by 3 gauges. If at least 2 gauges show the same pressure, then this value can be accepted as the true one. (But for safety reasons the failed gauge has to be replaced immediately.)

At a given time point $t_0$, subsystem $S_i$ is operating with probability $p_i$, $i = 1, 2, 3$. What is the probability $p_s$ that the system $S$ is operating at time point $t_0$?

Let $A_S$ be the event that $S$ is working at time point $t_0$, and $A_i$ be the event that $S_i$ is operating at time point $t_0$. Then,

$$A_S = (A_1 \cap A_2) \cup (A_1 \cap A_3) \cup (A_2 \cap A_3).$$

With $A = A_1 \cap A_2$, $B = A_1 \cap A_3$, and $C = A_2 \cap A_3$, formula (1.17) can be directly applied and yields the following representation of $A_S$:

$$P(A_S) = P(A_1 \cap A_2) + P(A_1 \cap A_3) + (A_2 \cap A_3) - 2P(A_1 \cap A_2 \cap A_3).$$

In view of the independence of the $A_1$, $A_2$, and $A_3$, this probability can be written as

$$P(A_S) = P(A_1)P(A_2) + P(A_1)P(A_3) + P(A_2)P(A_3) - 2P(A_1)P(A_2)P(A_3).$$

or

$$P(A_S) = p_1 p_2 + p_1 p_3 + p_2 p_3 - 2 p_1 p_2 p_3.$$

In particular, if $p = p_i$, $i = 1, 2, 3$, then

$$P(A_S) = (3 - 2p)p^2. \qquad \square$$

*Disjoint* and *independen*t random events are causally not connected. Nevertheless, sometimes there is confusion about their meaning and use. This may be due to the formal analogy between their properties:

If the random events $A_1, A_2, ..., A_n$ are disjoint, then, by formula (1.10),

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

If the random events $A_1, A_2, ..., A_n$ are independent, then, by formula (1.29),

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n).$$

## 1.5 EXERCISES

**Sections 1.1–1.3**

**1.1)** A random experiment consists of simultaneously flipping three coins.

(1) What is the corresponding sample space?

(2) Give the following events in terms of elementary events:
$A$ = '*head* appears at least two times,' $B$ = '*head* appears not more than once,' and $C$ = 'no *head* appears.'

(3) Characterize verbally the complementary events of $A$, $B$, and $C$.

**1.2)** A random experiment consists of flipping a die to the first appearance of a '6'. What is the corresponding sample space?

**1.3)** Castings are produced weighing either 1, 5, 10, or 20 *kg*. Let $A$, $B$, and $C$ be the events that a casting weighs 1 or 5*kg*, exactly 10*kg*, and at least 10*kg*, respectively. Characterize verbally the events $A \cap B$, $A \cup B$, $A \cap \overline{C}$, and $(\overline{A} \cup \overline{B}) \cap C$.

**1.4)** Three randomly chosen persons are to be tested for the presence of gene $g$. Three random events are introduced:

$A$ = 'none of them has gene $g$,'
$B$ = 'at least one of them has gene $g$,'
$C$ = 'not more than one of them has gene $g$'.

Determine the corresponding sample space $\Omega$ and characterize the events $A \cap B$, $B \cup \overline{C}$, and $\overline{B \cap \overline{C}}$ by elementary events.

**1.5)** Under which conditions are the following relations between events $A$ and $B$ true:
(1) $A \cap B = \mathbf{\Omega}$, (2) $A \cup B = \mathbf{\Omega}$, (3) $A \cup B = A \cap B$?

**1.6)** Visualize by a Venn diagram whether the following relations between random events $A$, $B$, and $C$ are true:
(1) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
(2) $(A \cap B) \cup (A \cap \overline{B}) = A$,
(3) $A \cup B = B \cup (A \cap \overline{B})$.

**1.7)** (1) Verify by a Venn diagram that for three random events $A$, $B$, and $C$ the following relation is true: $(A \backslash B) \cap C = (A \cap C) \backslash (B \cap C)$.
(2) Is the relation $(A \cap B) \backslash C = (A \backslash C) \cap (B \backslash C)$ true as well?

**1.8)** The random events $A$ and $B$ belong to a σ–algebra $\boldsymbol{E}$.
What other events, generated by $A$ and $B$, must belong to $\boldsymbol{E}$ (see definition 1.2)?

**1.9)** Two dice $D_1$ and $D_2$ are simultaneously thrown. The respective outcomes of $D_1$ and $D_2$ are $\omega_1$ and $\omega_2$. Thus, the sample space is $\mathbf{\Omega} = \{(\omega_1, \omega_2);\ \omega_1, \omega_2 = 1, 2, ..., 6\}$.
Let the events $A$, $B$, and $C$ be defined as follows:
$A$ = 'The outcome of $D_1$ is even and the outcome of $D_2$ is odd',
$B$ = "The outcomes of $D_1$ and $D_2$ are both even".
What is the smallest σ–algebra $\boldsymbol{E}$ generated by $A$ and $B$ ('smallest' with regard to the number of elements in $\boldsymbol{E}$)?

**1.10)** Let $A$ and $B$ be two disjoint random events, $A \subset \mathbf{\Omega}$, $B \subset \mathbf{\Omega}$.
Check whether the set of events $\{A, B, A \cap \overline{B}$, and $\overline{A} \cap B\}$ is (1) an exhaustive and (2) a disjoint set of events (Venn diagram).

**1.11)** A coin is flipped 5 times in a row. What is the probability of the event $A$ that 'head' appears at least 3 times one after the other?

**1.12)** A die is thrown. Let $A = \{1, 2, 3\}$ and $B = \{3, 4, 6\}$ be two random events.
Determine the probabilities $P(A \cup B)$, $P(A \cap B)$, and $P(B \backslash A)$.

**1.13)** A die is thrown 3 times. Determine the probability of the event $A$ that the resulting sequence of three integers is strictly increasing.

**1.14)** Two dice are thrown simultaneously. Let $(\omega_1, \omega_2)$ be an outcome of this random experiment, $A$ ='$\omega_1 + \omega_2 \leq 10$' and $B$ = '$\omega_1 \cdot \omega_2 \geq 19$.'
Determine the probability $P(A \cap B)$.

**1.15)** What is the probability $p_3$ to get 3 numbers right with 1 ticket in the '6 out of 49' number lottery?

**1.16)** A sample of 300 students showed the following results with regard to physical fitness and body weight:

| | | weight [$kg$] | | |
|---|---|---|---|---|
| | | 60 < | [60-80] | 80 > |
| | good | 48 | 64 | 11 |
| fitness | satisfactory | 22 | 42 | 29 |
| | bad | 19 | 17 | 48 |

One student is randomly chosen. It happens to be *Paul.*

(1) What is the probability that the fitness of Paul is satisfactory?

(2) What is the probability that the weight of Paul is greater than 80 $kg$?

(3) What is the probability that the fitness of Paul is bad and that his weight is less than 60 $kg$?

**1.17)** Paul writes four letters and addresses the four accompanying envelopes. After having had a bottle of whisky, he puts the letters randomly into the envelopes. Determine the probabilities $p_k$ that $k$ letters are in the 'correct' envelopes, $k = 0, 1, 2, 3$.

**1.18)** A straight stick is broken at two randomly chosen positions. What is the probability that the resulting three parts of the stick allow the construction of a triangle?

**1.19)** Two hikers climb to the top of a mountain from different directions. Their arrival time points are between 9:00 and 10:00 a.m., and they stay on the top for 10 and 20 minutes, respectively. For each hiker, every time point between 9 and 10:00 has the same chance to be the arrival time. What is the probability that the hikers meet on the top?

**1.20)** A fence consists of horizontal and vertical wooden rods with a distance of 10 $cm$ between them (measured from the center of the rods). The rods have a circular sectional view with a diameter of 2$cm$. Thus, the arising squares have an edge length of 8$cm$. Children throw balls with a diameter of 5$cm$ horizontally at the fence. What is the probability that a ball passes the fence without touching the rods?

**1.21)** Determine the probability that the quadratic equation

$$x^2 + 2\sqrt{a}\,x = b - 1$$

does not have a real solution if the pair $(a,b)$ is randomly chosen from the quarter circle $\{(a, b);\ a, b > 0,\ a^2 + b^2 < 1\}$.

**1.22)** Let $A$ and $B$ be disjoint events with $P(A) = 0.3$ and $P(B) = 0.45$. Determine the probabilities $P(A \cup B)$, $P(\overline{A \cup B})$, $P(\overline{A} \cup \overline{B})$, and $P(\overline{A} \cap B)$.

**1.23)** Let $P(A \cap \overline{B}) = 0.3$ and $P(\overline{B}) = 0.6$. Determine $P(A \cup B)$.

**1.24)** Is it possible that for two events $A$ and $B$ with $P(A) = 0.4$ and $P(B) = 0.2$ the relation $P(A \cap B) = 0.3$ is true?

**1.25)** Check whether for 3 arbitrary random events $A$, $B$, and $C$ the following constellations of probabilities can be true:

(1) $P(A) = 0.6$, $P(A \cap B) = 0.2$, and $P(A \cap \overline{B}) = 0.5$,

(2) $P(A) = 0.6$, $P(B) = 0.4$, $P(A \cap B) = 0$, and $P(A \cap B \cap C) = 0.1$,

(3) $P(A \cup B \cup C) = 0.68$ and $P(A \cap B) = P(A \cap C) = 1$.

**1.26)** Show that for two arbitrary random events $A$ and $B$ the following inequalities are true: $P(A \cap B) \le P(A) \le P(A \cup B) \le P(A) + P(B)$.

**1.27)** Let $A$, $B$, and $C$ be 3 arbitrary random events.

(1) Express the event '$A$ occurs, but $B$ and $C$ do not occur' in terms of suitable relations between these events and their complements.

(2) Prove: the probability of the event 'exactly one of the events $A$, $B$, or $C$ occurs' is

$$P(A) + P(B) + P(C) - 2P(A \cap B) - 2P(A \cap C) - 2P(B \cap C) + 3P(A \cap B \cap C).$$

**Section 1.4**

**1.28)** Two dice are simultaneously thrown. The result is $(\omega_1, \omega_2)$. What is the probability $p$ of the event '$\omega_2 = 6$' on condition that '$\omega_1 + \omega_2 = 8$?'

**1.29)** Two dice are simultaneously thrown. By means of formula (1.24) determine the probability $p$ that the dice show the same number.

**1.30)** A publishing house offers a new book as standard or luxury edition and with or without a CD. The publisher analyzes the first 1000 orders:

| | | luxury edition | |
|---|---|---|---|
| | | yes | no |
| with CD | yes | 324 | 82 |
| | no | 48 | 546 |

Let $A$ ($B$) the random event that a book, randomly choosen from these 1000, is a luxury one (comes with a CD). (1) Determine the probabilities

$$P(A),\ P(B),\ P(A \cup B),\ P(A \cap B),\ P(A|B),\ P(B|A),\ P(A \cup B|\overline{B}),\text{ and } P(\overline{A}\,|\overline{B}).$$

(2) Are the events $A$ and $B$ independent?

**1.31)** A manufacturer equips its newly developed car of type *Treekill* optionally with or without a tracking device and with or without speed limitation technology and analyzes the first 1200 orders:

| | | speed limitation | |
|---|---|---|---|
| | | yes | no |
| tracking device | yes | 74 | 642 |
| | no | 48 | 436 |

Let $A$ ($B$) the random event that a car, randomly chosen from these 1200, has speed limitation (comes with a tracking device).

(1) Calculate the probabilities $P(A)$, $P(B)$, and $P(A \cap B)$ from the figures in the table.

(2) Based on the probabilities determined under a), only by using the rules developed in section 1.3.3, determine the probabilities

$$P(A \cup B),\ P(A|B),\ P(B|A),\ P(A \cup B|\overline{B}),\ \text{and}\, P(\overline{A}\,\big|\,\overline{B}).$$

**1.32)** A bowl contains $m$ white marbles and $n$ red marbles. A marble is taken randomly from the bowl and returned to the bowl together with $r$ marbles of the same color. This procedure continues to infinity.

(1) What is the probability that the second marble taken is red?

(2) What is the probability that the first marble taken is red on condition that the second marble taken is red? (This is a variant of *Pólya's urn problem*.)

**1.33)** A test procedure for diagnosing faults in circuits indicates no fault with probability 0.99 if the circuit is faultless. It indicates a fault with probability 0.90 if the circuit is faulty. Let the probability of a circuit to be faulty be 0.02.

(1) What is the probability that a circuit is faulty if the test procedure indicates a fault?

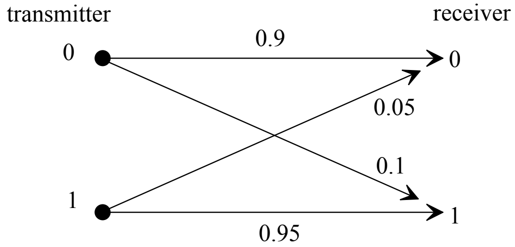
(2) What is the probability that a circuit is faultless if the test procedure indicates that it is faultless?

**1.34)** Suppose 2% of cotton fabric rolls and 3% of nylon fabric rolls contain flaws. Of the rolls used by a manufacturer, 70% are cotton and 30% are nylon.

a) What is the probability that a randomly selected roll used by the manufacturer contains flaws?

b) Given that a randomly selected roll used by the manufacturer does not contain flaws, what is the probability that it is a nylon fabric roll?

**1.35)** A group of 8 students arrives at an examination. Of these students 1 is very well prepared, 2 are well prepared, 3 are satisfactorily prepared, and 2 are insufficiently prepared. There is a total of 16 questions. A very well prepared student can answer all of them, a well prepared 12, a satisfactorily prepared 8, and an insuffi-




ciently prepared 4. Each student has to draw randomly 4 questions. Student *Fra*nk could answer all the 4 questions. What is the probability that Frank

(1) was very well prepared,
(2) was satisfactorily prepared,
(3) was insufficiently prepared?

**1.36)** Symbols 0 and 1 are transmitted independently from each other in proportion 1 : 4. Random noise may cause transmission failures: If a 0 was sent, then a 1 will arrive at the sink with probability 0.1. If a 1 was sent, then a 0 will arrive at the sink with probability 0.05 (figure).

(1) What is the probability that a received symbol is '1'?

(2) '1' has been received. What is the probability that '1' had been sent?

(3) '0' has been received. What is the probability that '1' had been sent?

**1.37)** The companies 1, 2, and 3 have 60, 80, and 100 employees with 45, 40, and 25 women, respectively. In every company, employees have the same chance to be retrenched. It is known that a woman had been retrenched (event $B$).

What is the probability that she had worked in company 1, 2, and 3, respectively?

**1.38)** John needs to take an examination, which is organized as follows: To each question 5 answers are given. But John knows the correct answer only with probability 0.6. Thus, with probability 0.4 he has to guess the right answer. In this case, John guesses the correct answer with probability 1/5 (that means, he chooses an answer by chance). What is the probability that John knew the answer to a question given that he did answer the question correctly?

**1.39)** A delivery of 25 parts is subject to a quality control according to the following scheme: A sample of size 5 is drawn (without replacement of drawn parts). If at least one part is faulty, then the delivery is rejected. If all 5 parts are o.k., then they are returned to the lot, and a sample of size 10 is randomly taken from the original 25 parts. The delivery is rejected if at least 1 part out of the 10 is faulty.

Determine the probabilities that a delivery is accepted on condition that

(1) the delivery contains 2 defective parts,

(2) the delivery contains 4 defective parts.

**1.40)** The random events $A_1, A_2, ..., A_n$ are assumed to be independent. Show that

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1-(1-P(A_1))(1-P(A_2))\cdots(1-P(A_n)).$$

**1.41)** $n$ hunters shoot at a target independently of each other, and each of them hits it with probability 0.8. Determine the smallest $n$ with property that the target is hit with probability 0.99 by at least one hunter.

**1.42)** Starting a car of type *Treekill* is successful with probability 0.6. What is the probability that the driver needs no more than 4 start trials to be able to leave?

**1.43)** Let $A$ and $B$ be two subintervals of [0, 1]. A point $x$ is randomly chosen from $[0, 1]$. Now $A$ and $B$ can be interpreted as random events, which occur if $x \in A$ or $x \in B$, respectively. Under which condition are $A$ and $B$ independent?

**1.44)** A tank is shot at by 3 independently acting anti-tank helicopters with one anti-tank missile each. Each missile hits the tank with probability 0.6. If the tank is hit by 1 missile, it is put out of action with probability 0.8. If the tank is hit by at least 2 missiles, it is put out of action with probability 1.

What is the probability that the tank is put out of action by this attack?

**1.45)** An aircraft is targeted by two independently acting ground-to-air missiles. Each missile hits the aircraft with probability 0.6 if these missiles are not being destroyed before. The aircraft will crash with probability 1 if being hit by at least one missile. On the other hand, the aircraft defends itself by firing one air-to-air missile each at the approaching ground-to-air missiles. The air-to-air missiles destroy their respective targets with probablity 0.5.

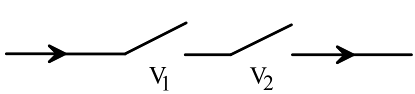
(1) What is the probability that $p$ the aircraft will crash as a result of this attack?

(2) What is the probability that the aircraft will crash if two independent air-to-air missiles are fired at each of the approaching ground-to-air-missiles?

**1.46)** The liquid flow in a pipe can be interrupted by two independent valves $V_1$ and $V_2$, which are connected in series (figure). For interrupting the liquid flow it is sufficient if one valve closes properly. The probability that an interruption is achieved when necessary is 0.98 for both valves. On the other hand, liquid flow is only possible if both valves are open. Switching from 'closed' to 'open' is successful with probability 0.99 for each of the valves.

(1) Determine the probability to be able to interrupt the liquid flow if necessary.

(2) What is the probability to be able to resume liquid flow if both valves are closed?



# CHAPTER 2

# One-Dimensional Random Variables

## 2.1 MOTIVATION AND TERMINOLOGY

Starting point of chapter 1 is a random experiment with sample space $\mathbf{\Omega}$, which is the set of all possible outcomes of the random experiment under consideration, and the set ($\sigma$–algebra) $\boldsymbol{E}$ of all random events, where a random event $A \in \boldsymbol{E}$ is a subset of the sample space: $A \subseteq \mathbf{\Omega}$. In this way, together with a probability function $P$ defined on $\boldsymbol{E}$, the probability space $[\mathbf{\Omega}, \boldsymbol{E}, P]$ is given. In many cases, the outcomes (elementary events) of random experiments are real numbers (throwing a die, counting the number of customers arriving per unit time at a service station, counting of wildlife in a specific area, total number of goals in a soccer match, or measurement of lifetimes of organisms and technical products). In these cases, the outcomes of a series of identical random experiments allow an immediate quantitative analysis. However, when the outcomes are not real numbers, i.e. $\mathbf{\Omega}$ is not a subset of the real axis or the whole real axis, then such an immediate numerical analysis is not possible. To overcome this problem, a real number $z$ is assigned to the outcome $\omega$ by a given real-valued function $g$ defined on $\mathbf{\Omega}$: $z = g(\omega)$, $\omega \in \mathbf{\Omega}$.

Examples for situations like that are:

1) When flipping a coin, the two possible outcomes are $\omega_1$ = '*head*' and $\omega_2$ ='*tail*'. A '1' is assigned to *head* and a '0' to *tail* (for instance).

2) An examination has the outcomes $\omega_1$ = '*with distinction*', $\omega_2$ ='*very good*', $\omega_3$ ='*good*', $\omega_4$ = 'satisfactory', and $\omega_5$ = 'not passed'. The figures '5', '4', $\cdots$, '1' (for instance) are assigned to these verbal evaluations.

3) Even if the outcomes are real numbers, you may be more interested in figures derived from these numbers. For instance, the outcome is the number $n$ of items you have produced during a workday. For first item you get a financial reward of \$10, for the second of \$11, for the third \$12, and so on. Then you are first of all interested in your total income per working day.

4) If the outcomes of random experiments are vectors of real numbers, it may be opportune to assign a real number to these vectors. For instance, if you throw four dice simultaneously, you get a vector with four components. If you win, when the total sum exceeds a certain amount, then you are not in the first place interested in the four individual results, but in their sum. In this way, you reduce the complexity of the ran- dom experiment.

5) The random experiment consists in testing the quality of 100 spare parts taken randomly from a delivery. A '1' is assigned to a spare part which meets the requirements,

and a '0' otherwise. The outcome of this experiment is a vector $\vec{\omega} = (\omega_1, \omega_2, \cdots, \omega_{100})$, the components $\omega_i$ of which are 0 or 1. Such a vector is not tractable, so you want to assign a summarizing quality parameter to it to get a random experiment, which has a one-dimensional result. This can be, e.g., the relative frequency of those items in the sample, which meet the requirements:

$$z = g(\vec{\omega}) = \frac{1}{100} \sum_{k=1}^{100} \omega_k \,. \tag{2.1}$$

Basically, application of a real function to the outcomes of a random experiment does not change the 'nature' of the random experiment, but simply replaces the 'old' sample space with a 'new' one, which is more suitable for the solution of directly interesting numerical problems. In the cases 1 and $3-5$ listed above:

1) The sample space $\{tail, head\}$ is replaced with $\{0, 1\}$.

3) The sample space $\{0, 1, 2, 3, 4, ...\}$ is replaced with $\{0, 10, 21, 33, 46, ...\}$.

4) The sample space $\{(\omega_1, \omega_2, \omega_3, \omega_4);\ \omega_i = 1, 2, ..., 6\}$, which consists of $6^4 = 1296$ elementary events of the structure $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)$, is replaced with the sample space $\{6, 7, ..., 24\}$.

5) The sample space consisting of the $2^{100}$ elementary events $\omega = (\omega_1, \omega_2, ..., \omega_{100})$ with $\omega_k$ is 0 or 1 is reduced by the relative frequency function $g$ given by (2.1) to a sample space with 101 elementary events:

$$\{0, \frac{1}{100}, \frac{2}{100}, \cdots, \frac{99}{100}, 1\}.$$

Since the outcome $\omega$ of a random experiment is not predictable, it is also random which value the function $g(\omega)$ will assume after the random experiment. Hence, functions on the sample space of a random experiment are called *random variables.* In the end, the concept of a random variable is only a somewhat more abstract formulation of the concept of a random experiment. But the terminology has changed: One says on the one hand that as a result of a random experiment *an elementary event has occurred*, and on the other hand, *a random variable has assumed a value*. In this book (apart from Chapter 12) only real-valued random variables are considered. As it is common in literature, random variables will be denoted by capital Latin letters, e.g. $X$, $Y$, $Z$ or by Greek letters as $\zeta$, $\xi$, $\eta$.

Let $X$ be a random variable: $X = X(\omega)$, $\omega \in \mathbf{\Omega}$. The *range* $R_X$ of $X$ is the set of all possible values $X$ can assume. Symbolically: $R_X = X(\mathbf{\Omega})$. The elements of $R_X$ are called the *realizations* of $X$ or their *values*. If there is no doubt about the underlying random variable, the range is simply denoted as $R$.

*A random variable $X$ is a real function on the sample space $\mathbf{\Omega}$ of a random experiment. This function generates a new random experiment, whose sample space is given by the range $R_X$ of $X$. The probabilistic structure of the new random experiment is determined by the probabilistic structure of the original one.*

When discussing random variables, the original, application-oriented random experiment will play no explicit role anymore. Thus, a random variable can be considered to be an abstract formulation of a random experiment. With this in mind, the probability that $X$ assumes a value out of a set $A$, $A \subseteq R$, is an equivalent formulation for the probability that the random event $A$ occurs, i.e.

$$P(A) = P(X \in A) = P(\omega, X(\omega) \in A).$$

For one-dimensional random variables $X$, it is sufficient to know the *interval probabilities* $P(I) = P(X \in I)$ for all intervals: $I = [a, b)$, $a < b$, i.e.

$$P(X \in I) = P(a < X \le b) = P(\omega, a < X(\omega) \le b). \tag{2.2}$$

If $R$ is a finite or countably infinite set, then $I = [a, b)$ is simply the set of all those realizations of $X$, which belong to $I$.

**Definition 2.1** The *probability distribution* or simply *distribution* of a one-dimensional random variable $X$ is given by a rule $\boldsymbol{P}$, which assigns to every interval of the real axis $I = [a, b]$, $a < X \le b$, the probabilities (2.2). ●

*Remark* In view of definition 1.2, the probability distribution of any random variable $X$ should provide probabilities $P(X \in A)$ for any random event $A$ from the sigma algebra $\boldsymbol{E}$ of the underlying measurable space $[\boldsymbol{\Omega}, \boldsymbol{E}]$, i.e. not only for intervals. This is indeed the case, since from measure theory it is known that a probability function, defined on all intervals, also provides probabilities for all those events, which can be generated by finite or countably infinite unions and conjunctions of intervals. For this reason, a random variable is called a *measurable function* with regard to $[\boldsymbol{\Omega}, \boldsymbol{E}]$. This application-oriented text does not explicitely refer to this measure-theoretic background and is presented without measure-theoretic terminology.

> *A random variable X is fully characterized by its range $R_X$ and by its probability distribution. If a random variable is multidimensional, i.e. its values are n-dimensional vectors, then the definition of its probability distribution is done by assigning probabilities to rectangles for $n = 2$ and to rectangular parallelepipeds for $n = 3$ and so on.*

In chapter 2, only one-dimensional random variables will be considered, i.e., their values are scalars.

The set of all possible values $R_X$, which a random variable $X$ can assume, only plays a minor role compared to its probability distribution. In most cases, this set is determined by the respective applications; in other cases there prevails a certain arbitrariness. For instance, the faces of a die can be numbered from 7 to 12; a 3 (2) can be assigned to an operating (nonoperating) system instead of 1 or 0. Thus, the most important thing is to find the probability distribution of a random variable.

Fortunately, the probability distribution of a random variable $X$ is fully characterized by one function, called its (*cumulative*) *distribution function* or its *probability distribution function*:

**Definition 2.2** The *probability distribution function* (*cumulative distribution function* or simply *distribution function*) $F(x)$ of a random variable $X$ is defined as

$$F(x) = P(X \le x), \quad -\infty \le x \le +\infty. \qquad \bullet$$

Any distribution function $F(x)$ has the following obvious properties:

1) $F(-\infty) = 0, \; F(+\infty) = 1,$ (2.3)

2) $F(x_1) \le F(x_2)$ if $x_1 \le x_2$. (2.4)

On the other hand, every function $F(x)$ satisfying the conditions (2.3) and (2.4) and being continuous on the left can be considered the distribution function of a random variable.

Given the distribution function of $X$, it must be possible to determine the interval probabilities (2.2). This can be done as follows:

For $a < b$, the event "$X \le b$" is given by the union of two disjoint events:

$$"X \le b" = "X \le a" \cup "a < X \le b".$$

Hence, by formula (1.11), $P(X \le b) = P(X \le a) + P(a < X \le b)$, or, equivalently,

$$P(a < X \le b) = F(b) - F(a). \tag{2.5}$$

Thus, the cumulative distribution function contains all the information, specified in definition 2.1, about the probability distribution of a random variable. Note that definition 2.2 refers both to discrete and continuous random variables:

> *A random variable $X$ is called* discrete *if it can assume only finite or countably infinite many values, i.e., its range $R$ is a finite or a countably infinite set. A random variable $X$ is called* continuous *if it can assume all values from the whole real axis, a real half-axis, or at least from a finite interval of the real axis or unions of finite intervals.*

Examples for discrete random variables are:

Number of flipping a coin to the first appearance of 'head', number of customers arriving at a service station per hour, number of served customers at service station per hour, number of traffic accidents in a specified area per day, number of staff being on sick leave a day, number of rhinos poached in the Krüger National park a year, number of exam questions correctly answered by a student, number of sperling errors in this chapter.

Examples for continuous random variables are:

Length of a chess match, service time of a customer at a service station, lifetimes of biological and technical systems, repair time of a failed machine, amount of rainfall per day at a measurement point, measurement errors, sulfur dioxide content of the air (with regard to time and location), daily stock market fluctuations.

## 2.2 DISCRETE RANDOM VARIABLES

### 2.2.1 Probability Distribution and Distribution Parameters

Let $X$ be a discrete random variable with range $R = \{x_0, x_1, \cdots\}$. The probability distribution of $X$ is given by a *probability mass function* $f(x)$. This function assigns to each realization of $X$ its probability $p_i = f(x_i)$; $i = 0, 1, \ldots$. Without loss of generality it can be assumed that each $p_i$ is positive. Otherwise, an $x_i$ with $f(x_i) = 0$ could be deleted from $R$. Let $A_i =$ "$X = x_i$" be the random event that $X$ assumes value $x_i$. The $A_i$ are mutually disjoint events, since $X$ cannot assume two different realizations at the same time. The union of all $A_i$,

$$\bigcup_{i=0}^{\infty} A_i,$$

is the certain event $\mathbf{\Omega}$, since $X$ must take on any of its realizations. (A random experiment must have an outcome.) Taking into account (1.9), a probability mass function $f(x)$ has two characteristic properties:

$$1)\ f(x_i) > 0, \quad 2)\ \Sigma_{i=0}^{\infty} f(x_i) = 1. \tag{2.6}$$

Every function $f(x)$ having these two properties can be considered to be the probability mass function of a discrete random variable. By means of $f(x)$, the probability distribution function of $X$, defined by (2.3), can be written as follows:

$$F(x) = \begin{cases} 0 & \text{if } x < x_0, \\ \sum\limits_{\{x_i, x_i \le x\}} f(x_i) & \text{if } x_0 \le x. \end{cases}$$

With $p_i = f(x_i)$, an equivalent representation of $F(x)$ is

$$F(x) = P(X \le x) = \begin{cases} 0 & \text{for} \quad x < x_0, \\ \Sigma_{i=0}^{k} p_i & \text{for} \quad x_k \le x < x_{k+1}, \quad k = 0, 1, 2, \cdots. \end{cases}$$

Figure 2.1 shows the typical graph of the distribution function of a discrete random variable $X$ in terms of the *cumulative probabilities* $s_i$:

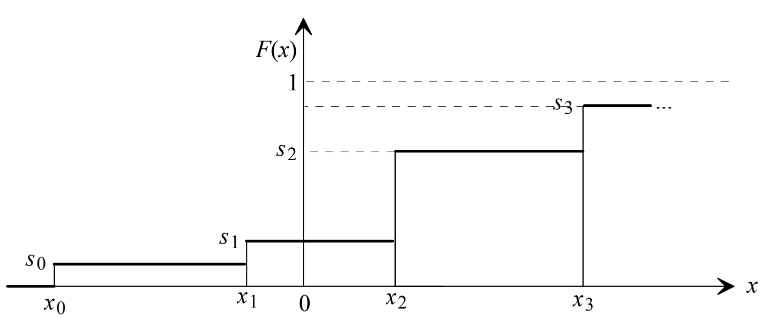



Figure 2.1 Graph of the distribution function of an arbitrary discrete random variable