E-LEARNING PROGRAM STOCHASTKION MAGISTER

www.stochastikon.com

COURSE 1:

MODELING UNCERTAINTY

Xiaomin Zhai & Elart von Collani
Stochastikon GmbH
Schießhausstr. 15, D-97072 Würzburg, Germany

Contents

Preface: Emergence of Science **v**

Course 1: Modeling Uncertainty **1**

Module 1.1: Basic Contents 2
Unit 1.1.1: Evolution 3
Unit 1.1.2: Connectivity 7
Unit 1.1.3: Uncertainty 10
Unit 1.1.4: Prediction 14
Unit 1.1.5: Science 17
Unit 1.1.6: Bernoulli Stochastics 26
Unit 1.1.7: Variability 30
Unit 1.1.8: Randomness 32
Unit 1.1.9: Probability 35
Unit 1.1.10: Ignorance 40
Unit 1.1.11: Knowledge 43
Module 1.2: Quantification 48
Unit 1.2.1: Features and Variables 50
Unit 1.2.2: Random Variables 53
Unit 1.2.3: Deterministic Variable 55
Unit 1.2.4: Ignorance Space 58
Unit 1.2.5: Variability Function 60
Unit 1.2.6: Random Structure Function 63
Unit 1.2.7: Probability Function 66
Unit 1.2.8: The Moments of a Probability Distribution 70
Unit 1.2.9: The Stochastic Model 75
Unit 1.2.10: Stochastic Science 82
Module 1.3: The Bernoulli Space 87
Unit 1.3.1: The Deterministic Case 88
Unit 1.3.2: The Case Without Ignorance 93
Unit 1.3.3: The General Case 98
Unit 1.3.4: Uncertainty Space 102
Unit 1.3.5: Representation of Probability Distributions 105
Unit 1.3.6: Families of Probability Distributions 107
Unit 1.3.7: Generic Deterministic Variable 111
Unit 1.3.8: Selection of Probability Distributions 113

Bibliography **120**

Index **122**

Preface: Emergence of Science

First of all a word of warning: The word Stochastics as used in this e-Learning Program was originally coined by Jakob Bernoulli in his masterpiece Ars conjectandi. It has nothing and really nothing in common with the word stochastics as being used in nowadays financial and the mathematical areas. Therefore, the name Bernoulli Stochastics is used for the content of this e-Learning Program. The aim of *Stochastikon Magister* is to make a contribution to the *Emergence of a Science*. This aim must be surprising, as science seems to have emerged centuries ago. In order to solve this puzzle, a short glance at contemporary science and philosophy of science will be beneficial. To this end a citation from an article authored by John R. Searle[1] is used, which was published in the 1999 issue of the Philosophical Transaction of the Royal Society, London.

The Philosophy of Science

In the 20th century, not surprisingly, the philosophy of science shared the epistemic obsession with the rest of philosophy. The chief questions in the philosophy of science, at least for the first half of the century, had to do with the nature of scientific verification, and much effort was devoted to overcoming various sceptical paradoxes, such as traditional problem of induction. Throughout most of the 20th century the philosophy of science was conditioned by the belief in the distinction between analytical and synthetic propositions. The standard conception of the philosophy of science was that scientists aimed to get synthetic contingent truths in the form of universal scientific laws. These laws stated very general truths about the nature of reality, and the chief issue in the philosophy of science had to do with the nature of their testing and verification. The prevailing orthodoxy, as it developed in the middle decades of the century, was that science proceeded by something called the hypothetico-Deductive method. The scientists formed the hypothesis, deduced logical consequences from it, and then tested those consequences in the form of experiments. This conception was articulated, I think more or less independently, by Karl Popper[2] and Carl Gustav Hempel[3].

Those practising scientists who took an interest in the philosophy of science at all, tended, I think, to admire Popper's views, but much of their admiration was based on a misunderstanding. What I think they admired in Popper was the idea that science proceeds by acts of originality and imagination. The scientist has to form a hypothesis on the basis of his own imagination and guess-work. There is no scientific method for arriving at hypotheses. The procedure of the scientist is then to test the hypothesis by performing experiments and reject those hypotheses that have been refuted.

Most scientists do not, I think, realize how anti-scientific Popper's views actually are. On Popper's conception of science and the activity of scientists, science is not an accumulation of truths about nature, and the scientist does not arrive at truths about nature, rather, all that we have in sciences are a series of so far unrefuted hypotheses. But the idea that the scientist aims after truth, and that in various sciences we actually have an accumulation

[1]John R. Searle, Slusser Professor of Philosophy, University of California - Berkeley.
[2]Karl Raimund Popper, Austro-British philosopher and a professor at the London School of Economics.
[3]Carl Gustav Hempel, German-American philosopher of science and a major figure in 20th-century logical empiricism.

of truths, which I think is the presupposition of most actual scientific research, is not something which is consistent with Popper's conception.

The comfortable orthodoxy of science as an accumulation of truths, or even as a gradual progression through the accumulation of so far unrefuted hypotheses, was challenged by the publication of Thomas Kuhn's[4] *'Structure of scientific revolutions' in 1962. It is puzzling that Kuhn's book should have had the dramatic effect that it did, because it is not strictly speaking about philosophy of science, but about the history of science. Kuhn argues that if you look at the actual history of science, you discover that it is not a gradual progressive accumulation of knowledge about the world, but that science is subject to periodic massive revolutions, where entire world views are overthrown when an existing paradigm is overthrown by a new scientific paradigm. It is characteristic of Kuhn's book that he implies, though as far as I know he does not state explicitly, that the scientist does not give us truths about the world, but gives us a series of ways of solving puzzles, a series of ways of dealing with puzzling problems within a paradigm. And when the paradigm reaches puzzles that it cannot solve, it is overthrown and a new paradigm is erected in its place, which again sets off a new round of puzzle-solving activities. From the point of view of this discussion, the interesting thing about Kuhn's book is that he seems to imply that we are not getting progressively closer to truth about nature in the natural sciences, we are just getting a series of puzzle-solving mechanism. The scientist essentially moves from one paradigm to another, for reasons that have nothing to do with giving an accurate description of an independently existing natural reality, but rather for reasons that are in greater or lesser degree irrational. Kuhn's book was not much welcomed by practising scientists, but it had an enormous effect on several humanities disciplines, especially those connected with the study of literature, because it seemed to argue that science gives us no more truth about the real world than do works of literary fiction or literary criticism; that science is essentially an irrational operation where groups of scientists form theories which are more or less arbitrary social constructs, and then abandon these in favour of other theories, which are likewise arbitrary social constructs.*

Whatever Kuhn's intentions, I believe that his effect on general culture, though not on the practice of real scientists, has been unfortunate, because it has served to demythologize science to debunk it, to prove that it is not what ordinary people have supposed it to be. Kuhn paved the way for the even more radical sceptical view of Paul Feyerabend[5]*, who argued that as far as giving us truths about the world is concerned, science is no better than witchcraft.*

My own view is that these issues are entirely peripheral to what we ought to be worried about in the philosophy of science, and what I hope we will dedicate our efforts to in the 21st century. I think the essential problem is this: 20th-century science has radically challenged a set of very pervasive, powerful philosophical and common sense assumptions about nature, and we simply have not digested the results of these scientific advances. I am thinking especially of quantum mechanics. I think that we can absorb relativity theory more or less comfortably because it can be construed as an extension of our traditional Newtonian conception of the world. We simply have to revise our ideas of space and time, and their relation to such fundamental physical constants as the speed of light. But quantum mechanics really does provide a basic challenge to our world view, and we simply have not yet digested it. I regard it as a scandal that philosophers of science, including physicists with an interest in philosophy of science, have not so far given us a coherent account of how quantum mechanics fits into our overall conception of the universe, particularly as regards to causation and determinacy.

[4]Thomas Samuel Kuhn, American physicist who wrote extensively on the history of science.

[5]Paul Karl Feyerabend, Austrian-born philosopher of science best known for his work as a professor of philosophy at the University of California, Berkeley.

> *Most philosophers, like most educated people today, have a conception of causation that is a mixture of common sense and Newtonian mechanics. Philosophers tend to suppose that causal relations are always instances of strict deterministic laws, and that cause and effect relation stand to each other in the kind of simple mechanical relations of gear wheels moving other gear wheels, and other such Newtonian phenomena. We know at some abstract level that is not right, but we still have not replaced our commonsense conception with a more sophisticated scientific conception. I think that the most exciting task of the 21st-century philosophy of science, and this is something for both scientists and philosophers, would be to give an account of the results of quantum mechanics that will enable us to assimilate quantum mechanics to a coherent overall world view. I think that in the course of this project we are going to have to revise certain crucial notions, such as the notion of causation; and this revision is going to have very important effects on other questions, such as the questions concerning determinism and free will. This work has already begun, and I hope it will continue successfully in the 21st century.*

Paul Feyerabend wrote in his book Against Method, (London, 1975 revised ed. London, 1988, p. 295):

> *science is much closer to myth than a scientific philosophy is prepared to admit. It is one of the many forms of thought that have been developed by man, and not necessarily the best. It is conspicuous, noisy, and impudent, but it is inherently superior only for those who have already decided in favour of a certain ideology, or who have accepted it without ever having examined its advantages and its limits.*

According to Popper's philosophical view, Kuhn's historical account and Feyerabend's scepticism, science does not and cannot provide any truths about reality. Scientific theories are imaginations and guesswork and scientific activities consists of solving puzzles. There is no scientific method and, as a consequence, whatever scientists do is called scientific. Note that Searle does not provide a single argument for his optimistic view on quantum mechanics, except that neither philosophers nor scientists have so far digested it. Thus, one may conclude that Feyerabend's view of science cannot be excluded and that science, in fact, is not what ordinary people have supposed it to be.

Even Searle admits that quantum mechanics indicates the collapse of the foundation of Modern Science which built on causation and determinacy and the belief in the truths being the result of the imagination of geniuses. The difficulty with quantum physics or radioactive decay is that it cannot be dealt with in a usually assumed cause-effect relation. Thus, science is suddenly based on two different world-views, which contradict each other.

The problem with searching truth is that truth assumes complete knowledge, but even geniuses like Aristotle and Newton had only partial knowledge or more precisely said were characterized by ignorance. Therefore, one needs a subjective imagination for filling the gap yielding hypotheses which although ingenious do not represent truth, which in fact is inaccessible for mankind. Another main problem is that a seemingly identified cause-effect relation necessarily appears in the shape of truth not admitting any room for expressing the always-existing uncertainty.

From the very beginning contemporary science claimed to be based on observation, i. e., experiments. However, the basic knowledge represented by the Natural Laws is - as described above - the result of the combination of some experiences and imagination. Experiments are mainly used for verifying the hypotheses and here again we meet a serious problem. So far none of the Natural Laws has ever been verified in the sense that a predicted event obtained by applying a Natural Law did really occur. Consequently, all the Natural Laws should be refuted. However, instead of refuting the hypotheses, science introduced the observation error and concluded that not the theory but the observation is wrong.

The e-learning Program *Stochastikon Magister* as part of the comprehensive *Stochastikon Information System* aims at introducing and teaching a new science based on a completely different world view and a different way of thinking than any of the traditional Western sciences. The new science is named Stochastic Science and it is founded on the here introsuced Bernoulli Stochastics. *Stochastic Science* does not search for or accumulate truths, but aims at gradually reducing ignorance, which is a characteristic feature of mankind.

For a better understanding of the contents of *Stochastikon Magister* a very brief account of the development of Science in Europe shall be provided below.

Scholastic Science

During the 9th century a movement of great significance was initiated in Europe. Schools were founded and the Holy Scripture lost its ultimate monopoly for wisdom. Besides the Bible the writings of the Church Fathers and of the Greek philosophers were studied and discussed. Public disputations pro and contra (Pierre Abelard: 'Sic et non') certain opinions became popular and led to the emergence of what is known as Scholastic Science. Scholastic Science was based

- on the application of logic,
- on clear definitions of questions and concepts,
- on the argumentation from effect to cause, and
- on the proof by authorities.

The assumption of God's omnipotence and omniscience led to the conclusion that God knows past as well as future and, hence, to a deterministic world view and to the confirmedness in cause-effect relations. It was believed that the Holy Scripture containing God's revelation enabled mankind to identify the causes and, thus, to make appropriate decisions.

Scholastic Science served Christian religion by verifying the 'truths' of the Holy Scripture and harmonizing 'rationality' and 'belief.' Still at the end of the 16th century, science was controlled by religion. The postulate of (Christian) religion truths was unquestioned and implied that

- Religion represents 'truth.'
- Science must be in accordance with 'truth.'
- Science aims at developing a rational foundation of 'truth.'

Greek, Arabian and Jewish advancements in philosophy, mathematics, astronomy and medicine were studied and gradually changed the minds of scholars and subsequently the world view in Europe.

Contemporary Science

During the 16th and 17th century something so far incredible happened. By observing nature and making experiments evidence was obtained which questioned the 'truths' of the Bible. Gradually the Holy Scriptures were replaced by the divine 'Book of Nature' by which God had designed the universe. Human ingenuity began to decipher chapter by chapter the Holy Book of Nature and detected the divine 'Laws of Nature'. These 'Laws of Nature' became the core of 'scientific theory' and it is widely believed that they represent 'truths'. The verification of these truths was performed by controlled experiments, and gradually the scientists became the authorities replacing the Church Fathers and Greek philosophers of Scholastic Science.

This development is called 'Scientific Revolution' and the result is often named 'Modern Science'. It may be characterized by the following issues:

- The search for 'truths,' called 'Natural Laws' coins 'Modern Science.'
- The 'Natural Laws' are formulated for non-existing 'ideal systems.'

- The belief in 'Natural Laws' is not really different from the belief in 'religious truths.'
- The verification of 'Natural Laws' is similar weird as the verification of divine laws used to be, as none of the 'Natural Laws' could ever be verified by observations.

The assumption that the evolution of universe can be separated in single cause-effect chains made it appear reasonable to investigate each 'cause-effect relation' individually as an isolated system. Consequently, evolution was divided into numerous narrow fields and for each field experts emerged for performing an analysis not being able for a holistic view.

The major differences between Scholastic Science and Modern Science are as follows:

- The disputation is replaced by experiments.
- The application of logic in the form of mathematics could be intensified by the advances of mathematics.
- The seemingly deciphering of the 'Book of Nature' made man believe to 'be able to master the nature.
- The 'Natural Laws' stimulated ingenious engineers to developed advanced technical devices which revolutionized the life of mankind.

The last point is often taken as justification of 'science.' However, if an engineer would actually use the Natural Laws, as formulated in the theory, for designing a device, it would not work. Therefore, engineers have to perform extensive experiments and by means of the trial-an-error method finally arrive at a feasible solution. Moreover, the advanced technology is developed without considering that in the long-run many new inventions turned out to be extremely dangerous for the vital systems, which are the basis of man's existence.

In summary, there is no essential difference between Scholastic Science' and Modern Science'. Both search for 'truths' in the form of cause-effect chains and, thus, are entirely based on the belief in a deterministic world, and exactly this is the problem. Evolution follows laws, but not deterministic ones.

Stochastic Science

At the end of the 17th century, Jakob Bernoulli, had the idea of a science of prediction, which he named Stochastics. His aim was to get rid of authorities and belief and to provide mankind with a scientific method to deal with uncertainty. Unfortunately, he passed away before he could establish the new science.

Stochastic Science is characterized by the following features:

- It aims at making reliable and accurate predictions about future developments.
- Predictions are made based on the verified knowledge about 'what is not', instead of the assumed knowledge about 'what is.'
- The process of acquiring new knowledge is essentially based on exclusion methods about 'what is not.'

In Stochastic Science there will be no paradigm changes, but a gradually decrease of ignorance about the laws which make evolution run. However, these laws will never be identified by mankind. The so far accumulated knowledge about nature must be relieved from determinism and reformulated. Thus, Stochastic Science can revise, refine and improve the knowledge and arrive at a more realistic description of evolution.

Stochastikon Magister aims at making a contribution to the Emergence of Stochastic Science, which is not based on causation and does not exhaust itself by solving puzzles and by such activities wasting enormous financial and human resources. *Stochastikon Magister* intends to introduce Bernoulli Stochastics in an easy-to-understand way, which, however, assumes the willingness to abandon the traditional way of thinking in favour of something which could be called *stochastic thinking*. This e-Learning Program is mainly based on two seminal papers by von Collani[6] and a booklet entitled "Stochastics" which was published in 2005 in China[7]

State of Stochastikon Magister

Finally, a word to the state of *Stochastikon Magister*: Bernoulli Stochastics is in a state of development and the same holds, of course, for this e-learning Program *Stochastikon Magister*. This second version is opened to public discussion in order to test its functionality and its comprehensibility. At the moment, only this English version is elaborated and should be used. We are working on the German translations and hope that soon also a complete German version is available. Of course, we would appreciate any proposal for improvement and any support in developing this e-learning program.

[6] von Collani (2004): "Theoretical Stochastics" and "Empircal Stochastics".

[7] von Collani & Zhai (2005): "Stochastics" (in Chinese).

Course 1: Modeling Uncertainty

Content, Aim and Benefits of the Course *Modeling Uncertainty*

Uncertainty about the future development is the most challenging issue mankind is facing at. If there would be no uncertainty, there would be no problems and, particularly, no need for 'science.' Uncertainty is not restricted to some areas of human activities, but is omnipresent. This implies that whatever man is doing uncertainty represents the main problem and, therefore, should be taken into account appropriately Thus, the ability of taking into account uncertainty is the main factor for success. Taking into account uncertainty appropriately assumes that a picture of what is called uncertainty is available.

The first course Modeling Uncertainty of this Learning Program in Bernoulli Stochastics consists of three modules. The first one introduces the basic concepts, which in the traditional sciences are kept unclear and confused. The second module takes up the concepts prepared in the first module and contains the rules how to quantify those quantities which are related to uncertainty. The last module adds the different pieces to one unique picture (model) of uncertainty.

The course offers the possibility for posing questions and taking part in tests. Frequent Asked Questions (FAQ) are displayed together with the answers. Moreover, a discussion forum may be used by students and teaching staff for joint communications. In all, Stochastikon Magister is designed in a way that it provides a virtual classroom with almost all possibilities offered by traditional classroom teaching. It is to be hoped that the participants will take advantage of the built-in functions.

Primary aim of the Course *Modeling Uncertainty* is to teach stochastic thinking, i. e., thinking in stochastic relations and abandon thinking in cause/effect relations. Because, training in causal thinking starts immediately after the birth of human beings, it becomes the 'natural way' of thinking and changing this habit proves to be extremely difficult.

The second aim of this course is to teach the rules how to model (describe) reality taking into account uncertainty. Because ignorance is the major source of uncertainty, it is of utmost importance to explicitly include it into the description.

A successful participation in this course, should result in a different perception of world, in a different evaluation of activities and, finally, in the desire to adopting a different way of making decisions.

Module 1.1: Basic Concepts

Content, Aim and Benefits of the Module *Basic Concepts*

Uncertainty about the further evolution of universe constitutes the center of stochastic science. Therefore, the fundamental concepts *evolution* and *universe* have to be explained as well as the ambivalent concepts of *past* and *future*. The basic question about what takes care of the astonishing order of the universe has to be answered as well as the issue how mankind fits into nature and how chaos comes in.

Finally more technical terms like *predictions*, *probability* and human *ignorance* versus human *knowledge* have to be treated.

Although uncertainty about the future development is certainly the most important issue of any human community, it has been more or less totally neglected by classical science. Probably this is the reason, why modern people have enormous difficulties to understand the stochastic concept of uncertainty. It demands to give up the idea of causality and this seems to be impossible by a majority of persons of our times.

Module 1.1 aims at preparing the ground for a change in thinking by clarifying some ambiguous concepts. This clarification is necessary, because otherwise the quantification of the concepts in the second module cannot be understood.

A successful passing of this module should lead to a better understanding of the universe and its continuously proceeding evolution.

In some sense the ideas, which are developed in this first module of *Stochastikon Magister*, seem to be rather obvious and more or less trivial. The fact that they have not been considered so far shows that humanity has not found its proper place within evolution, but floats without clear direction from one man-made calamity to the other. Especially when looking at the decision-making processes, it gets highly visible that mankind is ruled by egocentricity and particularist tendencies, which under the guise of secrecy unfold.

Unfortunately, none of the many religions which were developed, brought about a change. Also science, which is often looked upon as a guide in the dark, has been of no help. Religion and science themselves are controversial, maybe because both of them are essentially based on authorities and beliefs constituting a rather unreliable foundation. This basic module wants to present some thought-provoking impulses, which in the following modules are elaborated and exemplified.

Unit 1.1.1: Evolution

TARGET

The main goal of Learning Unit 1.1.1 is to impart the knowledge that future (= *not completed evolution*) is fundamentally different from past (= *completed evolution*). The second goal consists of showing that mankind's knowledge about the *evolution* is extremely limited and that other creatures have possibly a better understanding of evolution than human beings.

CONTENT

Universe, Nature and Evolution

The general meaning of the words *universe*, *nature* and *evolution* are as follows: In cosmology, the term universe (Latin: *universus* = entire) is understood as the entire space-time continuum including everything. The word nature is derived from the Latin word natura with the meaning of emerging or to be born. Nature is often considered as opposition to culture, in the sense that nature includes everything in the universe, which is not created by men. Evolution is derived from the Latin word *evolvere* which means 'to develop.' It is used for the development process by which living organism acquire and pass on novel characteristics.

In the framework of the e-learning Program *Stochastikon-Magister*, the term universe is used for the existing whole including everything and the term evolution is used for the transformation process the universe is permanently passing through. Finally, nature is used as a synonym of universe just as the word reality or real world.

What is the difference between Past and Future?

For mankind evolution becomes manifest by a continuously advancing change. The completed evolution is called *past*, while the not completed evolution is called *future*. Past is the realm of determinate facts, while future is the domain of indeterminate events. Determinate facts and indeterminate events are fundamentally different, as facts exist while future events refer to something which does not exist so far and which might never come into existence.

What is the difference between Facts and Events?

The actual set of facts represents one certain state of universe. Evolution means change of state, i. e., evolution is a process including the whole universe and producing one state after the other and turning indeterminate *events* into determinate *facts*. Facts may be known or unknown and future events may be anticipated.

Although there are seemingly many recurrent states from our experience it can be excluded that evolution contains iteration loops, i. e., each state of universe is different from all the states achieved in the past and all the states which will be achieved in the future.

What is Time

By comparing the flow of changes with certain seemingly recurrent events, mankind has introduced the concept of *time* and by this generated a certain order into the elapsing evolution.

For instance the rotation of earth which becomes manifest by the daily sunrise and sunset is used to define days, or the earth revolution which becomes manifest by the yearly seasons which is used to define a year. As a matter of fact, human beings have no direct sense for the artificially introduced concept time since the same time interval may seem to pass very slow or very fast according to the given situation. Thus, man needs clocks or certain recurrent events themselves as tools for orientation within the proceeding evolution.

Figure 1: A clock measuring the elapsing time by certain recurrent processes.

What is Space

Universe consists of many objects, some of them are visible for human beings others not. Similar as in the case of time, man has defined a three dimensional *space* for comparing the outward appearance of different objects by introducing systems of three-dimensional dimensions. The corresponding coordinates enable the orientation within those parts of universe which from the human perspective are persistent, i. e., hardly subject to evolution.

Similar as in the case of time, human beings have no direct sense for space or distance and need complicated tools for evaluating quantities defined by the coordinates. Introducing coordinates is useful and it is therefore justified. However, it should be clear that reality is different!

What is the Relation between Evolution and Man?

Mankind is part of the evolution, but human beings have no inner organ foreseeing the evolution and telling what to do in a certain situation to remain in accordance with evolution. It seems that many other creatures possess such an organ. If this is the case then one could not exclude that they are in a certain sense closer to evolution than mankind.

The more a creature must learn in order to survive the less knowledge it has about the ongoing evolution and, obviously, plants and animals need to learn less than man in order to survive. Fortunately, evolution does not advance blindly, but follows rules and laws which by experience can be discovered and utilized.

Strategies for Survival

The fact that evolution does not blindly advance, but follows rules enable mankind to survive despite the lack of closeness to evolution. The rules refer to the relation between completed and not completed evolution, i. e., the relation between past and future. For improving the chance and the ability for surviving, huge amount of resources are invested in order to understand a little bit of the rules of evolution and, thus, compensate to a certain extent the lack of closeness to evolution. Humans teach the acquired knowledge about the relations to the following generations. From birth onwards human beings have to learn and to try to develop *strategies for surviving* in an environment which is full of uncertainties for them. In fact mankind has learnt to survive, however, there is one central human realm which neglects uncertainty so far namely science and the main aim of this E-Learning program is to show, how uncertainty can be dealt with scientifically.

Examples

1. **Space and Time**

   Karl Pearson notes in his book *The Grammar of Science*[8]:

   > *Space and time are not realities of the phenomenal world, but the modes under which we perceive things apart. They are not infinitely large nor infinitely divisible, but are essentially limited by the contents of our perception.*

2. **Murder on the Orient-Express**

   The Orient Express started its first journey from Paris to Istanbul in 1883 and since then it has captured the imagination of the world (see announcement at the left). The train has been the temporary home of aristocrats, royalty, spies, film stars and writers - as well as, of course a certain fictional Belgian detective. During one of the travels the following message was disseminated by ticker:

   Not far from Belgrade, the luxurious Orient Express train is halted by a snowstorm. An American millionaire is found dead in his compartment. The door was locked, the window was open. But there were no tracks in the snow.

Figure 2: Orient Express advertisement.

   - The identity of the killed millionaire as well as the identity of the murderer are facts, i.e., *determined*. While the identity of the millionaire is *known* that of the murderer is *unknown*, i.e., represents the existing ignorance which shall be removed by an investigation process.
   - The future investigation is part of the not completed evolution and, thus, the outcome is *uncertain*. The investigations itself will consist of a series of activities by Hercule Poirot, the murderer and the other persons concerned. The future activities are subjected to randomness and therefore indeterminate.
   - Based on our knowledge that Hercule Poirot is infallible, the detection of the murderer is, of course, *certain*, however, the necessary time and other details remain *uncertain*.

3. **The Bowl Filled with Chocolate Balls**

   Consider a bowl filled with chocolate balls. You plan to take a fistful of balls out of the bowl.

   - The number if balls in the bowl is fixed and not at all *uncertain* although it is possibly *unknown* to you.

[8]Karl Pearson, The Universe of Science, London 1938, p. 229.

- The number of balls you will obtain by taking the fistful balls is generally *uncertain*, as it is not fixed (it does not exist), but will be the result of a so far not completed process.
- Next consider that there are only three chocolate balls is the bowl, which you easily can take by a fistful. In this case the future event about the number of balls is completely determined by the fact that there are only three balls and you will take them all.

4. **Fishing in a Carp Pond**

You fish regularly on Sunday morning in a nearby carp pond to spend your leisure time and provide your family with a fish dish.

- The number of carps in the pond and their whereabouts is fixed when you start fishing and not at all *uncertain*. However, you do not know neither the number nor the whereabouts.
- Even if you would know the exact number of carps and their exact whereabouts in the pond, the number of fishes you will catch during the morning is generally *uncertain*, because it refers to a future event and, thus, is part of the not completed evolution.
- However, if during the week the water in the carp pond had been emptied for catching the carps and consequently, there is no carp in the pond any more, then this fact would determine completely your fishing result.

5. **Solar System**

In the solar system nothing seems to be subject of evolution. The revolution of planets can be calculated by means of time and position and the future seems to be exactly determined by the past.

- In contrast to the above expressed belief, it is a simple fact that neither future positions nor future times of the planets in the solar system can be calculated exactly, because these values do not exist so far and are not at all completely determined by the past constellations.
- Moreover, even past and present positions of a planet within the solar system which are fixed and thus *certain* cannot be assessed exactly, but only approximately by an interval, which excludes those values which cannot be the true one. This illustrates the fact that human knowledge refers to "what is not" and not at all to "what is".

6. **Migratory Birds**

Migratory birds leave in autumn the northern countries and fly to southern countries before temperatures become to cold and return when the winterly conditions draw to a close.

- The birds have no global system of weather observation. Nevertheless, they leave the northern countries and return in due time. Therefore, it seems that these birds have a sense for the ongoing evolution.
- Even with a global system of weather observation installed, it seems that mankind cannot compete with migratory birds as to long-term weather forecast. If this is correct then we may conclude that man is more remote to evolution than animals.

Unit 1.1.2: Connectivity

Target

The main goal of Learning Unit 1.1.2 is to impart the knowledge that assuming cause effect relations as the rules of evolution prevents a better understanding of evolution and, thus, prevents a barrier to learning how to master problems due to an undesired course of the evolution.

Content

What is the Meaning of Connectivity?

Driving power of evolution, i. e., the continuous change universe passes trough, is the universal connectivity of everything with everything which keeps the development running, limits the number of future possibilities and determines the rules of change, i. e. the rules of evolution.

The best-known relation which indicates the all-inclusive connectivity is Gravitation which led Newton formulate his well-known Law[9]:

> *Every particle in the universe attracts every other particle with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.*

Although, this *Law of Nature* cannot be verified in a strict sense, as a verification would assume that all the masses in universe are known, it allows to exclude the possibility of the existence of isolated objects or systems in the universe and leads to the conclusion that in universe, everything is connected with everything similar as in any organism.

Note that talking about connectivity or interrelationship between different objects makes sense only if it refers to future changes where the state of each object affects the change with respect to each other object.

What are the Consequences of the Universal Connectivity?

Evolution makes universe continuously change from one state to the other. Because of the connectivity there are no independent subsystems and any change is generated by everything which exists in universe. Knowing the *truth* about universe is tantamount of knowing the state of universe for a moment in each detail. However, note that even knowing the *truth* would be without much value without knowing the rules of evolution.

As to the laws of evolution, the universal connectivity yields the following exclusions:

- The possibility of meaningfully break down the course of evolution into separate cause-effect chains can be excluded, as no single cause exist within a totally interrelated system.
- The possibility that man will ever identify *truth* can be excluded, as this would assume that every detail within universe is known.

[9]Definition from online-encyclopedia Wikipedia

- The existence of deterministic universal *Laws of Nature* that describe precisely the evolution can be excluded, as such laws assume the existence of identical initial conditions, which is impossible.

Human Perception of the Universe

Modern Science is based on the belief that evolution can be described by means of a deterministic machine, which operates according to parallel proceeding cause-effect chains. If this were correct, then each chain link, i. e., each cause-effect relation, constitutes an isolated micro-system and *truth* could be decomposed in small parts each referring to a thus defined subsystem.

However reality is different, because of the universal connectivity. Nevertheless, mankind searches and searches for *truth* and identifies an unending sequence of non-existing cause-effect relations, thus, preventing a better understanding of evolution.

The problem with an universal connectivity consists of how to take in into account. Working with a system described by a cause-effect relation is extremely simple. Each cause yields exactly one effect and, therefore, any cause-effect relation can be simply described by a mathematical function. In a completely connected system any observed effect can be generated by a multitude of different initial situations and any initial situation might generate a multitude of different results, implying that the traditionally used way of modeling relations fails. Therefore, wishing to take into account the universal connectivity necessitates the development of new methods for describing and analyzing any process of interest.

Traditional Science and Education in the Light of the Universal Connectivity

Traditional science is based on the assumption of isolated systems and, unfortunately, contemporary educational systems take up this prevailing scientific view which is known as reductionism. The mathematician Hyman Levy described the deplorable state of education already in 1932[10]:

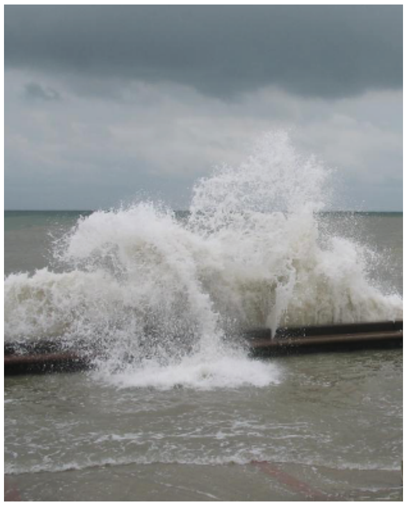
> *Man runs off easily along the tangent of speculation, isolating subjects and objects from their context, and building up elaborate structures on these isolated paths. From the changing matrix of the universe we separate out its biological, its chemical, its historical features. Our schools and our universities are designed to accentuate the isolation. Already at the school stage, and most certainly at the stages of higher education, experts in these "subjects" deal with their fields as if these existed by themselves, and as if their full significance could be derived from an internal study of these matters. So deep-rooted has specialist study become that the primary subject, what might perhaps be called Social Culture, of which these are mere subsidiary aspects, nowhere finds a place. The raw student emerging already highly specialised from school life, enters on the next stage of his career to submit to still more intensive specialisation. As a teacher he returns to school life to carry the process stage further. From a generation or two of this kind of practice there naturally emerges an elaborate philosophical justification embodied in such phrases as "Science for its own sake," as if the pursuit of anything could be a complete end in itself.*

One of the major aims of *Stochastikon Magister* is to point out that science should not be performed for its own sake and that the reductionist approach is the wrong one.

[10] Hyman Levy: *The Universe of Science.* Watts & Co., London, p. viii.

EXAMPLES

1. **Spring Tide**



The tide is the vertical rise and fall of the sea level surface. As the earth spins on its axis the centrifugal force results in slightly deeper water near the equator as opposed to shallower water at the poles generating a flow from the poles to the equator. The earth is also in orbit around the sun creating not only another centrifugal force but also a gravitational interaction. Moreover, the moon and the earth orbit each other creating also a centrifugal and a gravitational bulge. When sun and moon are aligned with the earth they combine their attraction and the combined effect of the sun and moon creates a spring tide.

Thus, it is known that whenever sun, moon and earth are aligned there will be a spring tide, defined as an unusually high rise of the sea level. However, this is not at all a cause-effect relation, as the same constellation of sun, moon and earth may lead to rather different values of the sea level on the one hand. Moreover, the same high sea level which may occur in a spring tide may also occur without sun, moon and earth being aligned.

2. **Health**

Similar as the universe, each part of the human body is connected with each part. Hearing bad news affects immediately the whole body and experiencing something joyful makes pain disappear. An influenza leads to problems with almost any part of the body and a ulcerated teeth may generate a great deal of different problems.

Although, the extremely close integration of the human body is well known, modern medicine considers and treats the human body similar as a deterministic machine and physicians search for cause-effect relations instead of regarding the entire organism which, of course, includes the mental, social and personal affairs, which probably affect the health in many instances much more than physical problems.

3. **Ecological Systems**

Recently, the trivial knowledge of connectivity has become popular in the realm of so-called ecological systems. An ecosystem can be regarded as an assemblage of organisms functioning within their specific environment as a connected unit where each part interacts with each part and any change or damage to one part affects the whole system.

Considering an ecological system as a whole and not as a number of isolated systems, is certainly an improvement. However, any ecological system is again only one part of a interacting larger system and disregarding this fact may lead to completely wrong ideas and decisions.

Unit 1.1.3: Uncertainty

TARGET

The main goal of Learning Unit 1.1.3 is to define the term "uncertainty" unambiguously and to distinguish clearly between *uncertainty* and its sources *ignorance* and *randomness*, which generate human uncertainty and which must be strictly distinguished and differently handled.

CONTENT

About the Meaning of Uncertainty

In everyday speech the word uncertainty has two different meanings:

- If uncertainty refers to a future development (i. e. to the not completed evolution) then it indicates that the development might take different courses leading to different results. This type of uncertainty is sometimes called "aleatory uncertainty".

- If uncertainty refers to a fact (i. e. to the completed evolution) then it indicates that the fact is not known. This type of uncertainty is sometimes called "epistemic uncertainty".

Whether a future event will occur or will not occur, cannot be known or unknown because it the decision will be made only in the future. The future can only be anticipated. In contrast to future events, facts exist and, therefore, cannot be anticipated, but are known or unknown. Anticipation with respect to future and ignorance with respect to past are two different things and, hence, speaking in both cases of uncertainty is mistakable. Therefore, in order to avoid confusion about the meaning, *uncertainty* is used in Bernoulli Stochastics exclusively with respect to *future events*. As to unknown facts the more appropriate word *ignorance* is used.

The Problem of Uncertainty

Mankind would have no problems, if there would be no uncertainty about future developments. Thus, the ultimate problem of mankind is uncertainty about future and, therefore, it is not surprising that a majority of human activities is devoted to overcome or at least reduce uncertainty.

The activities aim at establishing initial conditions which shall more or less guarantee an advantageous future. Such activities range from building up stocks to religious observance. The wanted conditions must be selected so as to be in accordance with the relations between past and future, i. e., in accordance with the rules of evolution. Only if these relations are known sufficiently well, it is possible to determine the conditions which reduce risks and uncertainty.

Unfortunately, man does not have a sense for identifying the relations which have to be taken into account, but must learn about them by own experiences or the experiences of others, which, however, assumes the possibility to communicate and circulate information. Thus, the language is a necessary means for reducing uncertainty by passing on knowledge.

Traditional Methods for Overcoming Uncertainty

Predictions are direct means for overcoming uncertainty. Thus, already ages and ages ago mankind developed divination for catching a glimpse on the future enabling appropriate decisions. However, a reliable prediction assumes knowledge about the actual situation on the one hand and knowledge about the rules of evolution on the other. If these preconditions are not met relying on a prediction made without this knowledge is hazardous.

Clearly, many animals have a better intuitive feeling for evolution and, thus, are able to act in many situations appropriately where man fails. Whether there are men, who have similarly an intuitive access to the otherwise hidden rules of evolution shall not discussed here. However, it is a fact that at all times swindlers took advantage of the belief in the practice of divination and, therefore, mankind developed a second method for overcoming uncertainty.

The much younger second human attempt to reduce uncertainty is called science. The development of its contemporary form started around the 16th century in Western Europe with astronomy. The great success of science in forecasting orbits of planets and stars became the basis of an unquestioned belief of man in science, which in some sense not only replaced to a wide extent divination but also religion.

Sources of Uncertainty

Obviously, human uncertainty about the future development can have only two source: one internal human source and one external source being independent of man. The internal source is called ignorance and refers to facts, i.e., the initial condition. The external source is a characteristic feature of evolution. It is called randomness and is the manifestation of the universal connectivity which excludes isolated systems and, hence, cause-effect relations. Everything is interrelated with everything and, therefore, no causes and effects exist, but an inherent variability in the outcome of all processes.

Randomness as the manifestation of universal connectivity yields order and, at the same time, variability. Thus, the changes produced by evolution are not static transformations, but dynamic changes. Ignorance, on the other hand results in seemingly disorder and chaos, which may lead to wrong decisions and produces the major problems mankind is facing. Presumably, there is no other creature which is more stuck with ignorance than man and, as mentioned before, ignorance is the main source of uncertainty about the future. Therefore, it is of utmost importance for mankind to develop appropriate rules how to deal with and how to reduce uncertainty, without the necessity to rely on subjective beliefs, which represent special manifestations of ignorance.

Figure 1 below illustrates the general situation - transition from past to future and the uncertainty involved. The figure shows knowledge or ignorance about the past on the left hand side by a set of four potential initial conditions each represented by a circle. The true but unknown initial condition is given by the black circle. The variability of the future development due to randomness is shown by the set of six possible outcomes, which are the only outcomes compatible with the true initial condition. The degrees of conformance between the true initial condition and the compatible outcomes are given by the corresponding probabilities represented by the horizontal bars set up in each of the future outcomes.

A prediction refers to a transition from past to future. Therefore, any mathematical model used for generating predictions should include the elements illustrated in this figure. These are

- the set of potential initial conditions,
- the set of possible future outcomes and

- the random structure over the set of possible outcomes.

The figure illustrates that starting with the initial condition reliably predicting a future event is possible. On the other hand, starting from the future outcome, it is also possible to determine a past fact. Therefore, the development of a set of rules how to describe and deal with uncertainty would be desirable.

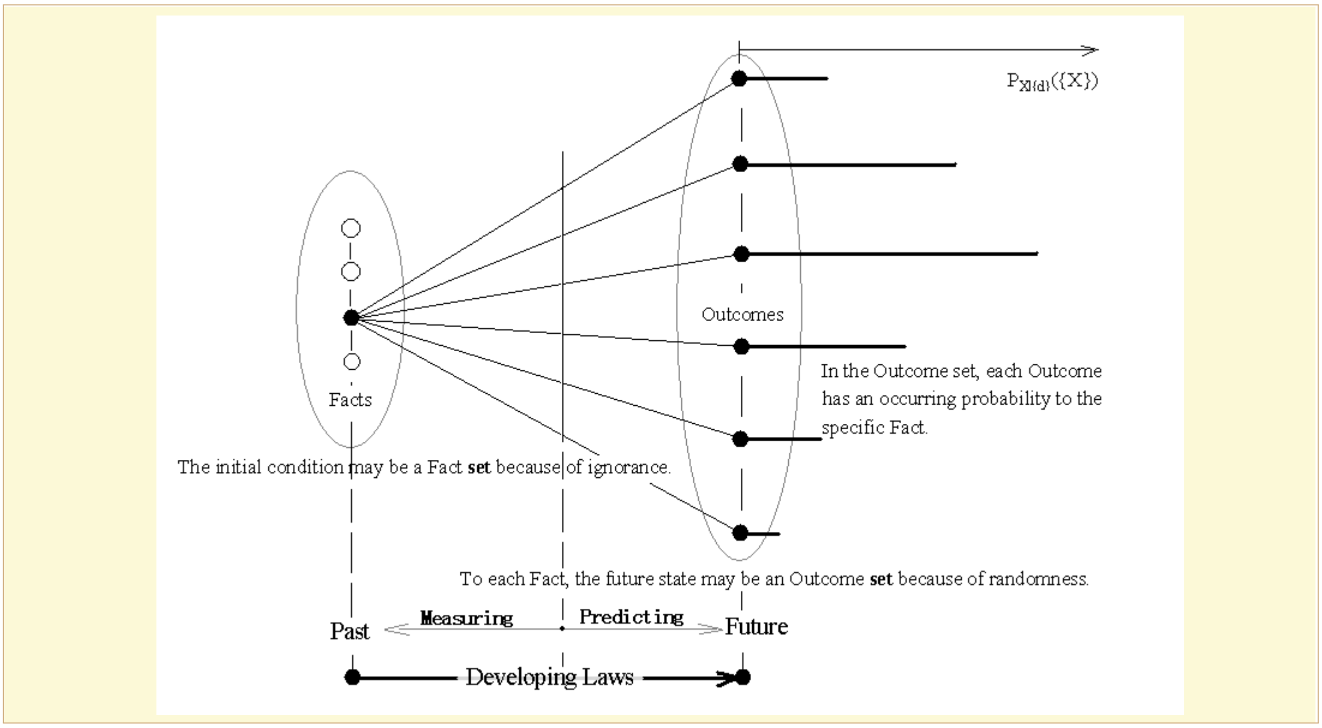


Figure 1: The general situation with respect to uncertainty - transition from past to future.

Rules for Dealing With Uncertainty

The rules for dealing with uncertainty must necessarily cover the following points:

- The quantities of interest of the future development. Identification of these quantities is of utmost importance, as a reliable prediction is only possible for a very limited part of the not completed evolution.
- The quantities of the past which are relevant for the quantity of interest in future.
- The existing ignorance as the internal source of human uncertainty.
- The structure of randomness as external source of uncertainty.

The rules should ensure that the resulting predictions are reliable implying that they must not be based on belief, but on sound knowledge. Sound knowledge on the other hand necessarily refers to *what is not*, as man will never be able to discover *what is*, i.e., the *truth*.

Examples

1. **Ignorance versus Uncertainty**

   The following examples shall illustrate the difference between uncertainty and ignorance.

- Murder on the Orient Express: Ignorance refers to the name of the murderer. Uncertainty refers to the future result of the investigation of the crime.
- Bowl filled with chocolate Balls: Ignorance refers to the (unknown) number of balls in the bowl. Uncertainty refers to the number of balls which will be drawn from the ball.
- Fishing in the carp pond: Ignorance refers to the number of carps in the pond. Uncertainty refers to the number of carps which will be caught during the Sunday morning.
- Solar system: Ignorance refers to the actual position of the planets and the sun. Uncertainty refers to any of their future positions.

2. **Sources of Uncertainty and Uncertainty**

The following examples shall illustrate the different sources which yield uncertainty.

- Bowl filled with chocolate balls: Assume that you can grab at least four and at most six chocolate ball with one fistful. Assume further that the bowl contains exactly three balls, but that this fact is unknown to you. In this case the complete uncertainty is generated by your ignorance.
- Next assume that you know that the bowl contains at least 20 balls. In this case the uncertainty about the number of balls you will draw is completely due to randomness. The ignorance about the exact number of chocolate balls does not increase your uncertainty.
- Fishing in a Carp Pond: Your uncertainty about the number of carps you will catch is due simultaneously to your ignorance about the number of carps in the pond and the randomness which rules your fishing.
- Solar system: The uncertainty of the future position of the planets in exactly 24 hours is due to your ignorance with respect to the actual positions and to randomness which yields certain variation in the orbits of the planets.

3. **Attempts to Overcome Uncertainty**

Humanity has developed many different methods for overcoming or at least reducing the human uncertainty about the future development.

- The canonical Chinese book "I Ching" the Book of Change describes a method for assessing the possible changes in future by means of throwing stalks of yarrow.
- Probably as old as the methods given in the I Ching are the methods developed in astrology, which use the constellations of planets and suns for overcoming uncertainty.
- Modern science restricts the predictions to very special processes which is described by a scientific theory, where the theory has been verified by controlled experiments.

Note that a majority of attempts to reduce uncertainty about the future development use experiments which cannot be controlled by man, for instance, random experiments, orbits of planets or putting a psychic into a trance. The results of such non-controllable experiments are not biased by human subjectivity and were thus supposed to reflect the universal connectivity which rules evolution.

Unit 1.1.4: Prediction

Target

Learning Unit 1.1.4 shall explain the concept of *prediction* since this term is often misunderstood leading to many errors and wrong decisions. The questions about the desirable properties of predictions and the necessary form of predictions are answered. In fact, in stochastic predictions play a central role in Bernoulli Stochastics reflecting the fact that the Greek word "stochastics" means "Science of Prediction" and consequently predictions constitute the heart of stochastics.

Content

Why Do We Need Good Predictions?

The problem is that one has to be prepared for what may occur in the future. To make an appropriate decision in a given situation it is necessary to know what may happen. Any statement saying what may happen in future is called a prediction. In case a prediction is available, the decisions can be made accordingly. However, whether a decision based on a prediction turns out to be appropriate depends on the quality of the prediction. A bad prediction can hardly lead to a good decision except in the case of luck, however, a good prediction may lead to an inappropriate decision, too. Hence, we conclude that a good prediction is a more or less necessary, but not a sufficient condition for making successful decisions.

What Is The Quality of a Prediction?

Generally, there will be a loss, if a decision is based on a prediction which does not occur. Thus the reliability of a prediction is of utmost importance.

Moreover, a prediction which is too vague, i.e., comprehends too many possible future events, also generates losses, as the decision has to take into account all the predicted events. Thus, the precision or accuracy of a prediction is another important feature.

We conclude that the quality of a prediction is determined by its reliability and its accuracy. Besides reliability and accuracy there is another important quality characteristic of predictions. Assume a reliable and accurate prediction, but it is communicated in a language which admits different interpretations. In such a case, only those will identify the true predicted event, who have found the correct interpretation. This leads to the demand that the prediction must be communicated in an unambiguous way that allows only a unique interpretation.

How Does a Good Prediction Look Like?

The required unambiguity of predictions demands to use an unequivocal language. The only unequivocal languages are formal languages with mathematics as the best developed one. It follows that a *good prediction* should be formulated in mathematical language. Mathematics uses numbers as letters and, therefore, quantification is obtained as a necessary requirement for making useful predictions.

Reliable predictions are the key for making appropriate decisions and, thus, for solving problems. However, a frequently made error is to assume that any good prediction must consist of one point only. Generally, the only founded statement about a one-point prediction is that it will not occur. The inherent variability due to randomness and the omnipresent human ignorance make it principally impossible to exactly foresee the indeterminate future. Therefore, a reliable prediction is in general a set of points that corresponds to a certain event. The higher the required reliability the larger is the set of points, i.e., the predicted event. Because mathematics is used for formulating a prediction, the predicted event is necessarily a set of numbers.

Ars Conjectandi

A *stochastic prediction* is given in the mathematical language. The predicted event is represented by a set, which is large enough to meet a specified reliability requirement with respect to its occurrence. The predicted event should be as small as possible, as the size of the set determines the accuracy of the prediction.

In 1713 Jakob Bernoulli's masterpiece Ars conjectandi was published eight years after Jakob had passed away. Jakob Bernoulli proposed to adopt *prediction* as the ultimate aim of scientific endeavors and to accept *probability* as the fundamental concept in science. Bernoulli had identified the *degree of certainty of the occurrence* as the characteristic feature of any future event. Subsequently he quantified the characteristic feature of future events by means of the concept of probability which constitutes the foundation of a science aiming at making reliable and precise predictions.

Examples

1. **Ambiguity of Predictions**

   A well-known example of an ambiguous prediction is the following: The Lydian king Croesus asked the Oracle of Delphy if he should attack the Persian king Cyrus. The Oracle made the prediction that if Croesus would cross the River Halys a great empire would fall. Based on the prediction and believing that the addressed empire would be Persia, Croesus attacked Persia, crossed the River Haly and was defeated by Cyrus.

2. **Unambiguity of Predictions**

   Apollo 11 was the first American manned spacecraft to land on the Moon. It launched from Florida on July 16, 1969, and four days later on July 20, mission commander Neil Armstrong and pilot Edwin 'Buzz' Aldrin became the first humans to set foot on the Moon.

   For each of the stages of the mission the control center had probably made prediction as time intervals. For instance the time to reach the lunar orbit could have been predicted by the following event:

   $$\{t \,|\, 43 \text{ h } 20 \text{ min} \leq t \leq 45 \text{ h } 15 \text{ min}\}$$

   Evidently, any prediction of such a form is unambiguous since the event is specified by numbers which have a unique interpretation.

3. **Reliability of Predictions**

   The reliability of a prediction refers primarily to the prediction generating process. This process yields reliable predictions if the predicted events generally occur.

   The reliability of a predicted event will decrease with decreasing size of the corresponding set. For instance, if in the example of the predicted arrival time for entering the lunar orbit the interval is shortened, then the reliability of the prediction gets probably smaller. If the interval shrinks to one point, then the predicted event will in general not occur with certainty.

4. **Precision of Predictions**

   If mathematics is used for formulating a prediction, then the involved quantities are quantified and the predicted event is necessarily a set of numbers with a specified size. The smaller the size of the set, the larger is the precision of the predictions. Thus, most precise prediction are singletons.

   A most precise prediction for entering the lunar orbit would be given, for example, by $\{44 \text{ h } 18 \text{ min } 33.345 \text{ sec}\}$. However, this time is more or less meaningless, as the only founded statement which can be made about this most precise prediction is that it will not occur.

Unit 1.1.5: Science

TARGET

Learning Unit 1.1.5 aims at making clear that *contemporary science* which is based on a deterministic and reductionist approach should be abandoned in favor of *stochastic science* representing a holistic approach that takes into account the inherent variability of all real processes. Students shall learn the weaknesses of contemporary science and, thus, understand the benefits of a change.

CONTENT

What is Science?

Since hundred of years the question *What is Science?* is discussed by philosophers (not so much by scientists) without arriving at a satisfactory answer. The inability of finding a solution indicates that there is something wrong, probably with both science and philosophy. In order to access meaning and significance of science one should first state its aim.

What is the Traditional Aim of Science?

Mankind dreams of omniscience and of acquiring the *truth*. Religion constitutes the traditional way of pretending to possess truth based on the assumption that there is an omniscient creature who promulgated truth to mankind, which makes predictions with respect to an eternal time possible.

When science started in Europe about 1000 years ago under the name scholasticism, its aim was defined as to verify the divine truths and thereby harmonize the Holy Scripture and reason. However, when it turned out about 500 years ago that the divine truths as given in the Bible were not consistent with reality, a new science evolved aiming at detecting the truth not in the Bible, but in the divine Book of Nature. Since then scientists have detected and promulgated endless *human truths*. Each of these truths has been refuted later or will be be refuted in future.

This is the actual state with respect to science. An incredible expense is made for obtaining something which is called “truth”, but which will be refuted with certainty in some future as it constitutes just a “guesses”, although often an ingenious one.

This last statement is acknowledged by scientist as for example by Michael Polanyi[11], who notes[12]:

> *The propositions of science thus appear to be in the nature of guesses. They are founded on the assumption of science concerning the structure of the universe and on evidence of observations collected by the methods of science; they are subject to a process of verification in the light of further observations according to the rules of science; but their conjectural character remains inherent in them.*

[11] Michael Polanyi, Hungarian-British polymath who worked in many areas as physical chemistry, economics, and philosophy.

[12] Michael Polanyi: *Science, Faith and Society.* The University of Chicago Press, Chicago, 1964, p. 31.

Does Contemporary Science Allow for Reliable Predictions?

Contemporary science emerged about 500 years ago in Europe. It constitutes an attempt of investigating all realms of universe for detecting the laws of evolution, which originally were thought to be designed by God in the mathematical language and which constitute the divine "Book of Nature". Since its emergence a large number of Natural Laws were invented, where a Natural Law explains reality by means of a causal relation and is believed to be universally valid. In order to verify the quality of Natural Laws one should compare the predicted events with the actually occurring ones. Such a comparison reveals that any prediction based on a Natural Law is wrong with almost certainty. Thus, relying on a Natural Law is tantamount (at least in general) with wrong decisions.

As a consequence when deciding on how to construct a device or executing a project, the Natural Laws must generally be modified by correction terms and factors which are determined by means of costly *trial and error*- experiments until the anticipated result is more or less obtained.

In fact, Natural Laws are not formulated for the real world, but for imaginary, ideal and isolated systems which do not exist. The Natural Laws are based on the assumption that evolution proceeds according to *cause-effect-relations* and, therefore, different parts of evolution may be considered as isolated processes.

That part of modern science which produces Natural Laws assuming a non-existing, ideal universe is generally called fundamental science and scientists consider it as the core of science. Thus, contemporary science cannot yield reliable predictions for any part of evolution, i. e., for any real world process and in fact, as explained above, contemporary science is not aiming at making reliable predictions.

The aim of contemporary science is to search and discover the truth which is believed to be fixed by nature or God in the form of deterministic Natural Laws. Thus, modern science and scholastic science are rather similar. Each aims at verifying an assumed divine truth. Scholastic science aims at verifying the truth as given in the Holy Scripture and expressed by certain religious authorities, modern science the truth as written in the Book of Nature in the form of mathematical Natural Laws and formulated by scientific authorities. In either case the methods for verifying are obscure and the used interpretations partly dictated by the aim. For example in modern science, any observation deviating from the truth does not lead to doubts in the so-called truth, but is declared to be an observation error. Believing in knowing the truth makes the consideration of uncertainty more or less unnecessary.

What Should be the Aim of Science?

Almost any problem of mankind is caused by uncertainty about the future development which makes appropriate decisions difficult or even impossible. As explained earlier reliable predictions are a promising way out of this difficulty. *Science* could be defined as the *Art of Making Reliable Predictions*. Following this idea, reliable predictions about future developments are set here to be the ultimate aim of science. One decisive prerequisite for arriving at reliable predictions should be mentioned already here. This is the exclusion of any subjective opinion or belief. If scientific results are based on subjective belief, no founded statement on the achieved reliability can be made.

The idea of an *Art of Prediction* was expressed 300 years ago by Jakob Bernoulli, who was professor for mathematics at Basel University. He coined the name *Ars conjectandi* or *Stochastics* for science, which means the art of making predictions.

The Question *What is Science?* Revisited

In ancient time divinationßindexdivination, astrology and other practices were used for making predictions. In our times the task has been overtaken by science, which investigates inanimate and animate nature in order to make predictions and supporting necessary decisions.

As a matter of fact both approaches failed at least from a more general point of view despite the enormous efforts made during all times and by all nations and communities. The failure of divination is obvious, the failure of science is concealed by a huge success of technology, but becomes visible, if one looks at the poor state of mankind and the miserable state of the vital systems (bio-diversity, air, water, soil, etc.) which are the basis of man.

However, more severe than the mere failure of science is the fact that training and education in schools and universities is based on the belief in determinism resulting in causal thinking and, hence, in an inappropriate handling of uncertainty. The consequences are wrong decisions and chaotic conditions in most fields of human activities.

Thus, a science aiming at coping with uncertainty and helping to make appropriate decisions has not emerged so far and it is high time for changing from an approach based on subjective belief (as represented by divination and contemporary science) to an approach which is based on objectively specifying what is not known. The latter approach is called *stochastic* science. It takes up the above stated ultimate aim of science, defines requirements for scientific products and develops the rules how to perform science. These rules make it possible to distinguish between *scientific* and *non-scientific* products, where science is to be understood as *stochastic science.*

The rules of stochastic science are developed and made available by Bernoulli Stochastics, which is the content of this e-learning Program *Stochastikon Magister.* Hence, this learning program is a means for the development of a science, which investigates real world as it presents itself and not as mankind believes it should be.

The Truth and its Societal Role

Possessing truth has always been a position of power and the origin of violence as long as mankind exists. The own truth had to be protected and had to be imposed upon those who believed in a different truth. Therefore, any pretended truth will lead almost inevitably to violence.

Religion, philosophy and also contemporary science represent the beliefs of certain groups of persons and, of course, each of the believers may totally rely on the respective belief. However, it should be clear that *belief* and *truth* are two extremely different concepts. *Belief* is necessarily bonded to a *human subject*, while *truth* refers to a fact and is necessarily bonded to the *evolution* and, particularly, independent of any human subject. In other words, any truth is universally valid, while a belief is valid only for the believer. Unfortunately, not only religion, but also contemporary science is based on belief. This fact may explain at least partially the strange and self-destructing thinking and acting of human beings.

Therefore, declaring any belief a truth constitutes a severe intervention into the personal spheres of other people, as already the declaration is an attempt to impose it upon others and, thereby, question the freedom of belief. If *freedom of belief* is included in the human rights, then calling a belief a truth must be regarded as a serious attack against human rights and, therefore, should be forbidden by law.

EXAMPLES

1. **Aim of Modern Science**

The aims of modern science are obscure and manifold depending on the individual who formulates them. Hyman Levy[13] quotes and comments[14] in his book "The Universe of Science" several individual opinions about the aims and problems of modern science:

- *"Our problem;" according to Professor A.N. Whitehead[15], "is to fit the world to our perceptions and not our perceptions to the world." Here Whitehead adopts the idealistic position. His own perceptions are to him prime reality and the world is a system fabricated to fit them.* (pp 195 and 196)
- *To Sir Arthur Eddington[16], on the other hand, "Science aims at constructing a world that shall be* symbolic *of the world of common-place experience." He does not consistently hold this view. He begins his* Nature of the Physical World *by drawing a distinction between the table familiar to him in everyday experience on the one hand, what Professor Whitehead would call the common-sense notion table – the table he write on – and an contra-distinction to this his scientific table whose vast emptiness is sparsely scattered with numerous electrons rushing about at great speeds, and "whose combined bulk amounts to less than a millionth part of the* table itself.*" Quite inconsistently, in spite of the "table itself," he adds later on, "Modern physics has by delicate test and remorseless logic assured me that my second scientific table is the only one that is* really *there," although what delicate tests these can be that involve the "thereness" of the scientific table without involving also the same "thereness" of the familiar table and of the testing apparatus he does not state.* (pp. 196 and 197.)
- *According to Mr. Bertrand Russell[17], on the other hand, "The aim of physics, consciously or unconsciously, has always been to discover what we may call the causal skeleton of the world." Elsewhere he says, "It is obvious that a man who can see knows things that a blind man cannot know; but a blind man can know the whole of physics. Thus the knowledge that other men have and he has not is not part of physics." (*the Analysis of Matter*, pp. 391 and 389.) If the knowledge that seeing people have is knowledge of the world of reality, and if it be true that a blind man may know the whole of physics, it seems evident that science cannot span the whole range of reality. It seems relevant to inquire, not whether sight can be dispensed with and yet leave the individual capable of knowing the whole of science, but which senses, if any, are essential for this purpose. If one sense at least is required for an individual to make contact with the world of reality, on what evidence can a distinction be drawn for this purpose between one sense and another? What Mr. Russell actually implies is that the scientific picture can be isolated from the world of sight, and that nothing that it offers is thereby lost. This is surely an unsubstantiated assertion. ... What in fact Mr. Russell has done has been to ignore the terms "of the world" in his description of the aim of physics "to discover the causal skeleton of the world." If such a bony structure is exposed, it is a mere disembodied skeleton, but an essential part of the make-up of the world, and cannot be isolated from it without loss.* (pp. 197 and 198)
- *In his* Grammar of Science*, Professor Karl Pearson[18] has stated that "the classification of facts, the recognition of their sequence and relative significance, is the function*

[13] Hyman Levy (1889-1975), Scottish mathematician and professor at the Imperial College of Science and Technology, London.

[14] Levy, (1938), pp. 195-200.

[15] Alfred North Whitehead (1861-1947), English mathematician and philosopher.

[16] Arthur Stanley Eddington (1882-1944), British astrophysicist.

[17] Bertrand Russell (1872-1970), British philosopher, logician, mathematician, and historian.

[18] Karl Pearson (1857-1936), established the discipline of mathematical statistics at University College, London.

of science," while the scientific attitude is shown in "the habit of forming judgement on these facts unbiased by personal feeling." "The scientific man," he says elsewhere. "has a strive at self-elimination in his judgements." In the same vein Dr. Dingle[19] *asserts in his* Science and Human Experience *that science is the "recording, augmentation, and rational correlation of those elements of our experience which are actually or potentially common to all normal people," and he goes on to amplify the meaning of this terms.* (p. 199)

- *"The object of all science." Professor Einstein*[20] *state in* The Meaning of Relativity, *"is whether natural science or psychology, is to co-ordinate our experiences and to bring them into a logical system." (p. 199)*
- *These are the dicta of men of science, theorists and practitioners alike. We may compare with them the point of view of a writer like J. W. Sullivan*[21]*, who derived æsthetic inspiration from music and mathematics. "The ideal aim of science." said that writer in* The Bases of Modern Science*, "is to give a complete mathematical description of phenomena in terms of the fewest principles and entities" (p. 22). Then again (p. 226): "The present tendency of physics is towards describing the universe in terms of mathematical relations between unimaginable entities."*

2. **Gravitational Law**

In 1687 Newton published his work on "The universal law of gravity" stating that that any two particles having masses m_1 and m_2 and being separated by a distance r are attracted by a gravitational force F given by:

$$F = g\frac{m_1\, m_2}{r^2} \tag{1}$$

where g is a "universal constant". Newton's "Natural Law" was refuted only in the last century and replaced by Einstein's "Natural Law of Gravity."

Newton's Law is taught worldwide in secondary schools and used for predicting the falling times of bodies. However, the teachers fail to explain that the "Law" does not hold on earth and that the predicted results are wrong with certainty, because the assumptions made are not met on earth and strictly speaking nowhere.

The philosopher Nancy Cartwright[22] comments the Law of Gravitation with the following words:

> *Does this Law truly describe how bodies behave?*
> *Assuredly not. Feynman*[23] *himself gives one reason why. 'Electricity also exerts forces inversely as the square of the distance, this time between charges . . .' It is not true that for any two bodies the forces between them are charged is given by the law of gravitation. Some bodies are charged bodies, and the force between them is not $g\frac{m_1 m_2}{r^2}$. Rather it is some resultant of this force with the electric forces to which Feynman refers.*
> *For bodies which are both massive and charged, the law of universal gravitation and the Coulomb's law (the law that gives the forces between two charges) interact to determine the final force. But neither law itself truly describes how*

[19] Herbert Dingle (1890-1978), British philosopher of science at University College London.
[20] Albert Einstein (1879-1955), German-American theoretical physicist and philosopher.
[21] John William Sullivan (1886-1937), popular science writer and literary journalist.
[22] Nancy Cartwright, born 1944, American philosopher of science.
[23] Richard Feynman (1918-1988), American physicist.

> *bodies behave. No charged objects will behave just as the law of universal gravity says; and any massive objects will constitute a counterexample to Coulomb's law. These two laws are not true; worse, they are not even approximately true. In the interaction between the electrons and the protons of an atom, for example, the Coulomb effect swamps the gravitational one, and the force that actually occurs is different from that described by the law of gravity.*
> *There is an obvious rejoinder: I have not given a complete statement of these laws, only a shorthand version. The Feynman version has an implicit ceteris paribus modifier in front, which I have suppressed. Speaking more carefully, the law of universal gravitational is something like this:*
>
>> *If there are no forces other than gravitational forces at work, then two bodies exert a force between each other which varies inversely as the square of the distance between them, and varies directly as the product of their masses.*
>
> *I will allow that this law is a true law, or at least one that is held true within a given theory. But it is not a very useful law.* [24]

Nancy Cartwrights words illustrate the following issues: Fundamental science formulates laws not for this world, but for an imaginary world, which is called 'ideal world'. The results achieved for the imaginary world are not only useless but also dangerous, because if applied might lead to a disaster. The reason for the failure of this type of science is the fact that the universal connectivity is neglected and, instead, non-existing isolated systems are considered. This becomes immediately obvious by the assumption "if there are no forces other than the gravitational forces at work" since such a condition does not hold in the entire universe.

As to Einstein's Law, it is not refuted so far, but with certainty it will be refuted in future.

3. Schrödinger Equation

In 1925 Erwin Schrödinger[25] described the time-dependence of quantum mechanical systems by means of a partial differential equation, which shall provide a quantitative description of the change of the (quantum) state of the system.

Just as in Newton's physics Schrödinger's partial differential equation does not allow for taking into account the universal connectivity and consequently the omnipresent variability of evolution. It assumes a universe which can be separated into small isolated systems which operate independently. Fortunately, this assumption is not met in universe as otherwise evolution would probably proceed chaotically.

Although, the quantum mechanical model given by the Schrödinger equation, does not include explicitly randomness, probabilities emerge by normalizing the wave function and then taking its absolute value resulting in a number between 0 and 1. This is interpreted[26] as the probability that the particle will be found in a certain location. This interpretation is more or less arbitrary and has been questioned since it was adopted.

[24] Cartwright, N. (1983), p. 57.
[25] Erwin Schrödinger (1887-1961), Austrian theoretical physicist.
[26] There were several proposals how to interpret the wave function, finally, the "probability interpretation" proposed by M. Born (1882–1970) was accepted.

4. **Engineering Science**

Besides fundamental science which, strictly speaking, deals and investigates a non-existing universe, there are the so-called engineering sciences, which develop technology. Engineers apply in a certain sense scientific knowledge obtained in fundamental science to solve technical problems. The crucial task of an engineer is to identify, understand, and integrate the constraints imposed by real world for producing an operating product. In other words, an engineer must try to correct the findings of fundamental science until reliable predictions are possible.

The general method for arriving at a viable prediction procedure are trial-and-error, i. e., an option is tried, if it works, a solution has been obtained. If it doesn't work, the option constitute an error and another option is tried. Trial and error aims not to improving understanding as it makes no attempt to discover why a solution works. Therefore, trial and error does not improve knowledge, but only yields an isolated solution to a specific problem.

The need for engineering science results from the fact that fundamental science is based on the untenable assumption that evolution consists of cause-effect chains, which makes the results of fundamental physics unfeasible. Nancy Cartwright writes about quantum physics and classical physics in her book *The Dappled World - A Study of the Boundaries of Science* as follows:

> *The conventional story of scientific progress tells us that quantum physics has replaced classical physics. We have discovered that classical physics is false and quantum physics is, if not true, a far better approximation to the truth. But we all know that quantum physics has in no way replaced classical physics. We use both; which of the two we choose from one occasion to another depends on the kinds of problems we are trying to solve and the kinds of techniques we are master of.*[27]

If the word "use" means to develop technical devices and not to illustrate scientific experiments, then Nancy Cartwright is wrong. Neither quantum physics nor classical physics can be used, but simple and costly trials have to be made until finally the desired result has been achieved. Unfortunately Cartwright's does not explain her claim that quantum mechanics is a better approximation to the truth than classical physics.

5. **Genesis of Human Brain Cells**

In Issue 20 of 2006 the title-story of the German newsmagazine *Der Spiegel* dealt with the topic how new cells develop in the brain and that it is hoped to find a way to cure nerves and brain by activating the development of new cells. On page 166 it is reported:

> *This hope is based on a phenomenon, which neuro-science during the last century has stubbornly denied. The opinion of the Spanish brain-researcher and later Nobel prize winner Ramón y Cayal was effective like a prohibition. In 1928 he had simply claimed 'In an adult brain the neurons are stable and unchangeable. Each of them can die but nothing can regenerate.'*
> *There were soon some doubts about the official doctrine, however, those experimenters, who expressed their doubts, where laughed at by their colleagues. Joseph Altmann of Massachusetts Institute of Technology (MIT) in Cambridge gave radioactive marked DNA material to adult rats, cats and guinea pigs.*

[27] Cartwright, 1999, p.2.

Subsequently he was able to detect the DNA material in neurones: In the course of the cell division they were built into the cell nucleus - a proof that neurones had newly developed in the brain. The scientific community, however, ignored Altmann's findings. MIT refused to give him a tenure track position – he found a position only in the remote Indiana and fall into oblivion.
Ten years later Michael Kaplan of the University of New Mexico presented photos of newly developed nerve cells taken with an electron microscope. But, he, too, was confronted with preconception. Kaplan remembers that the then influential brain-researcher Pasko Raskic of Yale University commented his findings by the words: 'The cells may look like neurones in New Mexico, but they do not in New Haven.'
Rakic even developed a theory why human nerve calls cannot divide: In some stage of the human developing process our ancestors had exchanged their ability to develop new neurones against the ability to store memories by maintaining a constant number of neurones. The brain of the Homo sapiens cannot develop new cells because of reasons of stability.'

This example illustrates the fact that scientific theories are in general not based on empirical observation, but, contrary, observations are neglected, because they contain evidence that the theory is wrong. In this special case, the consequences with respect to wrong diagnoses and wrong treatments can hardly be guessed. Moreover, the generations of highly talented researches misled by the belief in an authority constitute an incredible waste of human resources.

6. **Guardian of Order of the Universe**

Edmund Byrne[28] describes in [4] how man seeks truth for proving the order of universe. For him there is continuity in this aim throughout the different stages of the development of science up to or times:

Medieval thinkers, taking their cue from the Hellenic ancestors, were no less imbued with the love for order. Just as the Roman law smoothed out the unpleasant complexities of political life for the sake of what is still known as "the common good", so too did the Scholastics maintain order and discipline in matters intellectual. Greek cosmology remained important as a foundation for the medieval world-order, but became in fact but a visible sign of higher, spiritual order: the order of faith. Thus the guardian of order and regularity was no longer merely the political ruler nor even the religious leader, but the leader of the new intellectualism, the theologian, to whom fell the awesome task of describing and at times even creating order in a universe permeated with the divine.

Since that time, of course, a few changes have been made. The science of order par excellence, mathematics, has blossomed forth in marvellous profusion; and with the aid of the new mathematics undreamed of patters of regularity have been partly found, partly introduced into the physical universe. The mathematician – or, in the eyes of the masses, the physicists – has become the guardian of order in the universe. So powerful, indeed, has become the new mathematics, both in its pure and in its applied aspects, that one can no longer say readily where "science" leaves off and its object begins. The "universe", once somewhat naively looked upon as something to be discovered and explained, tends to be more and more the product of man's cogitation and creation. Thus, at least, did the great German philosopher Immanuel Kant envision the relationship between scientific

[28]Edmund F. Byrne (born 1933), American philosopher.

> *thought and its object. It is no longer au courant to refer to oneself as being Kantian; but few philosophers of science have been able to escape Kant's radical dichotomy between subject and object. The denouement of post-Newtonian absolutism, which was in fact a kind of naive mathematical realism, has led to what might be called, by comparison, mathematical idealism. The search for order is by no means less intense than in former days; quite the contrary. It is just that man is now much more conscious of the fact that the order of which he speaks is perhaps due as much to thought as to things; and the thought from whence that order arises is, as often as not, the thought of the mathematician.*[29]

Byrne describes how man does not look at the universe in order to describe it, but shapes it according to his own cogitation, which change from the Hellenistic an Roman times to the religious leaders of medieval times and finally arrived at mathematics and physics. But the aim remained the same, not to describe reality, but to explain it.

Reference

[1] Cartwright, N. (1983): *How the Laws of Physics Lie.* Cambridge, Clarendon Press, Oxford.

[2] Cartwright, N. (1999): *The Dappled World. A Study of the Boundaries of Science.* Cambridge, University Press.

[3] Levy, H. (1938): *The Universe of Science.* London.

[4] Byrne, E.F. (1968): *Probability and Opinion. A Study in the medieval Presuppositions of Post-Medieval Theories of Probability.* Martinus Nijhoff, The Hague.

[29] Byrne, from the author's preface, pp. XX, XXI.

Unit 1.1.6: Bernoulli Stochastics

TARGET

The aim of Learning Unit 1.1.6 is twofold:

1. The relation between mathematics and science shall be made clear, i.e., mathematics is solely a language with some superior properties when compared with natural languages. However, mathematics and reality have almost no relation to one another.
2. The rules developed in *Bernoulli Stochastics* and defining *stochastic science* shall be explained and illustrated.

CONTENT

Science and Mathematics

Before explaining the role of Bernoulli Stochastics within Science, we have to look at science and at mathematics in order to avoid common misunderstandings. One possibility to clarify the situation is to identify aim and constraints of science and mathematics and thus find differences and similarities.

- As explained in the previous learning unit science should aim at coping with uncertainty about future developments (evolution of universe) by detecting the constraints imposed by the past on the future development. Based on these findings, the future developments should then be predicted in an unambiguous, reliable and sufficiently accurate way.
- Mathematics aims at developing a system of notions and relations between numbers built on some abstract basic requirements (axioms). Any new proposition (statement of a relation) becomes part of mathematics, if it meets the constraints imposed by logic, where logic constitutes a set of rules how to infer new propositions from already established ones starting from the axioms.

Evidently the aims of science and mathematics have nothing in common. Science deals with real world and the relations between past and future, while mathematics deals with abstract entities and their formal (logic) relations. However, the requirement of unambiguity of scientific statements cannot be warranted unless mathematics is used as language. Therefore, mathematics constitutes a necessary means for science. Mathematics is based on numbers and this leads to an important precondition for using mathematics as scientific language: The situations investigated have to be quantified beforehand.

Mathematics is the scientific language but does not include any rules or directives how to perform science and in particular it contains no hints, which allow for explaining, describing or handling uncertainty. Therefore, starting with a mathematical concept and not with a real world entity leads often to unsolvable problems of interpretation.

Bernoulli Stochastics

Mathematics could not have been developed without a set of rules how to deduce propositions from previous ones. This set of rules is called logic. Similarly, science cannot develop without a set of rules how to describe and handle uncertainty. This set of rules is called Bernoulli Stochastics. The rules must answer the following questions

- What must the sources of uncertainty be quantified?
- How can uncertainty of future developments be mathematically described or modelled?
- How should a scientific statement be organized?
- Which requirements (specifications) should a scientific proposition meet?
- How shall the conformity with the specifications be checked?

Thus, the role of Bernoulli Stochastics is similar within science as logic is within mathematics. Just as logic enables to distinguish what is part of mathematics and what not, stochastics makes it possible to decide upon scientific and non-scientific statements. Bernoulli Stochastics provides the general directives and specifications for science.

Stochastic rules shall guarantee that:

- Scientific statements are unambiguous, as ambiguous statements lead to misunderstandings and wrong decisions.
- Scientific statements must not be based on assumptions or beliefs as otherwise science would reflect not real world but a subjective opinion.
- Any scientific statement is incomplete, if it is not accompanied by the degree of its reliability, since otherwise the risk of wrong decisions is unknown.
- The correctness of a scientific statement must be checkable.

Rules of Bernoulli Stochastics

Because in the real world everything is connected with everything, a model of the entire evolution is impossible and, therefore, should not be attempted. Below the stochastic rules are stated in a rather informal way:

- *Identify those aspects of the future development which are of interest. Quantify all the aspects of interest by assigning to each possible outcome a number according to the rules of quantification.*

Quantification of the future aspects of interest yields the corresponding variables. Because the future value of such a variable is indeterminate and depends on some random developments, these variables are called random variables denoted by a symbol, e.g., X.

Predictions are possible only because everything is connected with everything. This universal connectivity excludes a deterministic evolution in favor of a stochastic evolution. A scientific prediction is based on the relation between the future evolution and the actual, already completed state of evolution. For describing the relation between past and future, it makes sense to start with the actual state, i. e., the so-called initial conditions.

- *Identify all aspects of the initial conditions being relevant for the already identified random variables. Quantify all these aspects of the initial conditions by assigning to each potential value a number according to the rules of quantification.*

The result is a representation of the initial conditions by variables. The values of these variables are determinate and, therefore, the corresponding variables are called deterministic variables denoted by a symbol, e.g. D.

The pair of variables (X, D) represents the aspects of interest of the future and those aspects of the past which affect the future development with respect to X. Any scientific description of the considered process must start with the pair of variables (X, D). It should be clear that the same process may lead to many different pairs of variables depending on the interests at hand.

The initial conditions (represented by the deterministic variable D) affect the future development. Generally, the actual value of the deterministic variable D is unknown and, therefore, it cannot be taken as starting point for the desired prediction. But, there is no case thinkable, in which D can adopt any real number. It is always possible to exclude with certainty almost all numbers. Those numbers which cannot be excluded, constitute the set of potential values of D describing the existing ignorance and resulting in the next stochastic rule:

- *Identify all potential values of D by excluding all those values of D which are not compatible with the available knowledge about the initial conditions.*

The initial conditions affect the future development by ruling out almost all values for the specified random variables X. The set of remaining admissible values must be specified:

- *Identify for each of the values of D that cannot be excluded, a set of values which the random variable X may eventually adopt in future.*

The initial conditions not only limit the future developments to a certain set, but also attract each possible value of X with a different degree of strength. The initial conditions thereby define a structure on the set of possible future outcomes of X by assigning to each future event a probability or, equivalently, by distributing the probability mass on the possible future outcomes. The probability of a future event is the degree of certainty of its occurrence and, thus, reflects the strength of attraction emanating from the initial condition.

- *Identify for each set of potential initial conditions the corresponding probability distribution of the random variable.*

The above indicated rules describe the tasks that have to be solved in Bernoulli Stochastics. What is still needed are detailed instructions how to solve the corresponding problems, i.e., how to select, improve and verify the different set-functions needed to arrive at a complete stochastic model and to exploit and improve a given model.

Examples

1. **Geometry**

   Geometry is the first branch of mathematics which was build on axioms and logic. The first set of axioms was formulated in the *Elements* of Euclid of Alexandria (ca. 325 BC – 65 BC) where all theorems ("true statements") are derived from a finite number of axioms. Euclid gives five axioms:

   - Any two points can be joined by a straight line.
   - Any straight line segment can be extended indefinitely in a straight line.
   - Given any straight line segment, a circle can be drawn having the segment as radius and one endpoint as center.

- All right angles are congruent.
- Parallel postulate. If two lines are drawn which intersect a third in such a way that the sum of the inner angles on one side is less than two right angles, then the two lines inevitably must intersect each other on that side if extended far enough.

Note that the quantities like points, lines, angles are completely abstract quantities with not real meaning. Assigning a real meaning to them must be looked upon as an interpretation which might be justified or not.

Mathematics consist of deriving statements from the set of axioms using the rules of logic.

2. **Train Journey**

Consider travelling on 25 March 2006 by train from Würzburg Central Station to Frankfurt/Airport. The departure is schedule at 11.09 a.m. and the arrival 10.26 a.m. The problem is to predict the travel time in a reliable way.

- The aspect of interest of the future development is the travel time from Würzburg to Frankfurt. As the aspect of “time” has been quantified long ago, quantification consists of selecting an appropriate time-unit, say minutes, and the symbol to be used for the corresponding variable, say T.
- The relevant aspects of the given situation before the train leaves Würzburg determine the range of variability of T and the random structure on this range of variability. The relevant aspects and the corresponding variable must be selected appropriately aand are combined to the deterministic variable D.
- Next, those values for D which are not compatible with the given situation may be excluded resulting in a set of potential values of D.
- As soon as the potential initial conditions are available, the set of all those values for the travel time T which cannot be excluded on condition of the potential values of D must be derived.
- Finally the structure of randomness given by the probabilities of the possible outcomes must be determined,

After the situation before leaving Würzburg, has been described as outlined above, it becomes possible to make a prediction or answer any other question with respect to the travel time.

Unit 1.1.7: Variability

TARGET

Learning Unit 1.1.7 aims at introducing *variability* as the central concept for describing evolution and the state of human knowledge. The variability refers to the future value of any random variable, and it becomes manifest by the fact that for each repetition of a process the corresponding random variable will adopt a different value.

CONTENT

Randomness and Variability

Uncertainty refers to the future development and particularly to the future outcome of a specified random variable X, which at the time being is indeterminate, because of the universal connectivity of universe. Thus, the future development which will produce the outcome of interest may take the one or the other course. However, the universal connectivity does not only generate variability, it also limits variability. Thus, the connectivity provides something like a dynamic order and prevents chaos, which would be almost inevitable if the universe would consist of independent and isolated systems. The set of possible courses of development is generally rather small, but contains always more than one course and therefore more than one outcome of X. The actual selection of one of the possible courses or outcomes of X happens by random and, thus, randomness provides variability within evolution.

Note, that the above described variability cannot be excluded as a characteristic feature of evolution. Moreover, this variability is independent of human knowledge. Even if the truth (about the initial conditions) were exactly known, the variability would not vanish. This distinguishes randomness from ignorance. An almighty God is free of any ignorance, however, randomness as a characteristic property of the universe precludes the knowledge about the exact future outcome.

Ignorance and Variability

A decision should consider the possible future developments or more precisely said the variability in the future outcome of X. The variability is determined by the initial conditions, i. e., by facts. Because of the universal connectivity, mankind will never be able to know these facts completely. The available knowledge about the initial conditions is the result of excluding all impossible conditions yielding a set of potential conditions which includes the actual or true one. Any of the potential conditions might be the true one and, therefore, each of them, represented by a value of the deterministic variable D, has to be taken likewise into account. Thus, due to ignorance one has to take into account some variability with respect to the deterministic variable D although the latter is determinate. Clearly, the variability with respect to the deterministic variable D increases the variability which has to be considered with respect to the random variable X.

Variability and Knowledge

Uncertainty about a future development is expressed by variability and, therefore, variability should be the main object of interest in science. The variability of the deterministic variable D represents ignorance. Any diminishment of ignorance means a reduction of the variability of D or, equivalently an increased knowledge about the initial conditions, i.e., the past. Increasing the knowledge about the past conditions leads to a better understanding of the variability of X reflecting the dynamics of evolution. Thus, decreasing ignorance about the past makes better predictions about the future possible.

Examples

1. **Variability and Evolution**

   Gregor Johann Mendel (1822 – 1884) was an Augustinian monk who investigated the laws of inheritance by conducting plant hybridity experiments. The significance of Mendel's work was not recognized until the turn of the 20th century.

   His results are known as Mendel's Laws of Heredity and consists of statements about the variability exhibited by offsprings of given parents. The following figure shows the variability in the offsprings of a black and a white rabbit.

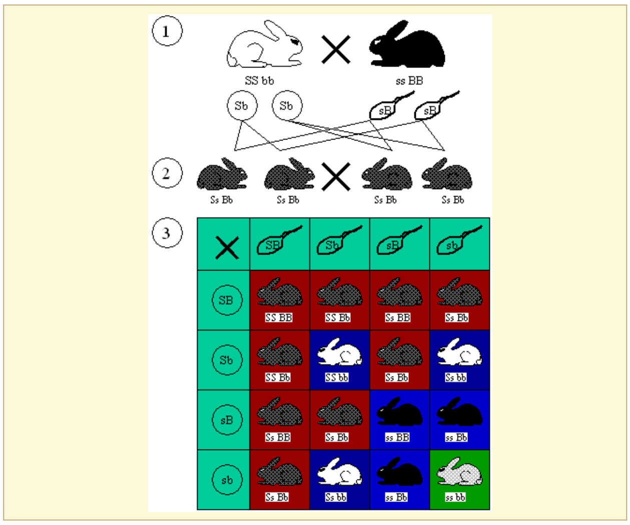


Figure 1[30]: Variability among the offsprings of a white and a black rabbit.

2. **Variability and Ignorance**

   Limited and structured variability is an inherent feature of evolution. In contrast, ignorance generates a seemingly variability which does not exist, but has to be taken into account when making predictions or decisions.

   Consider a car racing through the Western Sahara and let the number X of cars reaching successfully the goal be the quantity of interest. If the number n of participating cars is known, then the random variable X may adopt a value between 0 and n. However, if n is unknown and only an upper bound $n+m$ is known, then the range of variability for X which has to be taken into account increases to the set of integers between 0 and $n+m$.

[30]Figure taken from From Wikipedia, the free encyclopedia

Unit 1.1.8: Randomness

Target

Learning Unit 1.1.8 aims at introducing *randomness*, which represents the principle of order within evolution as it admits only few future developments and, moreover, gives priority to some of them. Randomness includes, for example, each of the natural forces considered in physics.

Content

What is Randomness?

Jacques Monod[31] believed *that the evolution of all living things is brought about by natural selection operating through entirely random variations. These are "the only possible source of modifications in the genetic text." Hence chance alone is at the source of every innovation, of all creatures in the biosphere." The concept that "pure chance, absolutely free but blind" is the basis of evolution, is now "the sole conceivable hypothesis, the only one compatible with observed and tested fact." Moreover, there is no warrant whatever for supposing the "conceptions about this sole, or ever could, be revised."*

In fact, the conception that chance or randomness acts "absolutely free but blind" is widespread. Probably it goes back to the quantification of games of chance in the famous correspondence between Blaise Pascal and Pierre de Fermat and the subsequent works by Christiaan Huygens, Pierre Rémond de Montmort and Abraham de Moivre, where chance is by construction "blind." Another source of this conception might be the belief in a deterministic world where the observed variability is caused by ignorance.

However, the number of scientists recognizing that randomness does not act "blindly" is increasing. Schilling[32] notes *Replying to Einstein's familiar question, Joseph Ford of the Georgia Institute of Technology declares, "God plays dice with the universe. But they're loaded dice. And the main objective of physics is to find out by what rules they were loaded, and how we can use them for our own ends."*

As explained in the previous learning units randomness as a characteristic feature of evolution is not at all free (of any structure) or even blind, but constitute the (only) universal principle of order.

Randomness as Manifestation of Connectivity

The past is characterized by facts which do not exhibit any variability, future exhibits variability and stands for evolution of universe. The available experience allows to exclude the existence of isolated systems, which yields a universal connectivity including inanimate and animate nature. The universal connectivity makes cause-effect relations impossible and does not – strictly speaking – allow for recurrent processes. Hence, future is characterized by variability

[31] Chance and Necessity, London 1972, cited according to S. Paul Schillling (1991): Chance and Order in Science and Theology. *Theology Today* 47, 1-11

[32] Paul Schillling (1991): Chance and Order in Science and Theology. *Theology Today* 47, 1-11

generated by the universal connectivity. As it is common practice, we say that the future development is subjected to randomness.

Randomness as manifestation of connectivity is the clear opposite of chaos which leads to disorder and confusion. In contrast randomness means not only limited, but also structured variability of the future development. The structure leads to the well-known fact that in similar situations (with respect to the initial conditions) some events occur more frequently than others.

Randomness the Principle of Order

As explained, randomness is generated by the universal connectivity and becomes manifest by variability of future developments. It may be quantified by the extent and the structure of variability. An unstructured variability would mean that each of the possible outcomes has the same chance of occurrence, while a structured variability means that some of the outcomes are preferred while others are so to say penalized. Some of the possible outcomes are attracted by the initial conditions more than others.

This latter fact reminds of attracting and repelling forces in physics. From a stochastic point of view each of the physical forces is a manifestation of universal connectivity and makes certain events more likely, others less likely and a vast majority of events impossible. Thus, the physical forces of nature can be looked upon as an important part of randomness as a characteristic feature of evolution.

Randomness and Gambling

During the 16th century Blaise Pascal and Pierre de Fermat started to investigate games of chance. This was the beginning of the development of a new branch of mathematics originally called *Aleae Geometria* by Blaise Pascal and which later became known as Probability Theory. There is an decisive difference between *chance* in the sense of gambling and *randomness* as it is introduced here. Randomness represents connectivity and, thus, order, while chance within gambling represents independence and, thus, disorder or chaos. Games of chance are designed by symmetries in a way that the range of variability is fixed, but no special structure is imposed on the future course of the game. Thus, chance in the sense of gambling can be looked upon as a man-made degeneration of randomness. This idea is also expressed in the above citation of Joseph Ford, where the “rules they were loaded” are the rules of evolution, i. e., the relations between facts and the future development.

Examples

1. **Fundamental Interactions in Physics**

   From physics it is known that particles interact with each other. Such interactions can be observed between galaxies as well as between quarks jiggling around inside a proton. The observed interactions enable to exclude the existence of isolated systems and necessitate the consideration of randomness as done in Bernoulli Stochastics.

   In physics four types of forces are identified gravity, electromagnetism, the weak interaction, and the strong interaction. In Bernoulli Stochastics these interactions are part of the overall interactions which altogether generate randomness.

   Physics has acknowledged that phenomena as the movements of galaxies, the development black holes, the expansion of the universe, the orbits of planets, the falling of objects, etc., etc are generated by these interactions, i. e., by randomness as expression of the universal connectivity. Nevertheless, physics insists of following reductionism which neglects the universal connectivity that is the driving force of evolution.

2. **Herd Instinct**

Besides the interactions within the inanimated nature, there are, of course, also interactions among creatures although they are not as well investigated as the physical interactions.

The "herding instinct" among human beings is an often observed tendency among man to identify themselves with a larger group of individuals and to copy behaviors and beliefs. The herding instinct of humans manifests in a number of ways, for example buying the same goods and performing the same activities, believing in the same doctrines and being of the same opinion.

In numerous cases the herding instinct among humans has been exploited and is exploited for prompting people to act against their own interests. The herding instinct shows that the existence of interactions of rather different types than the physical ones cannot be excluded and, therefore, have to be taken into account when modeling uncertainty.

3. **Interactions Among Animals**

Even more striking than the herding instinct among humans are observed interactions among animals. Ants are a species which exhibits extreme interactions between the individuals. They form highly organized colonies or nests and the ant colonies may be described as superorganisms because they appear to operate as a single entity. For instance, although there are millions of ants in a colony which frequent the narrow streets in a colony, traffic jams like in congested areas of humans cannot be observed.

Figure[33] 1: Ants Highway.

Clearly, any "social behavior" among creatures constitutes a form of connectivity which immediately changes the structure of randomness.

[33]Figure taken from From Wikipedia, the free encyclopedia: Ants forming a highway seen in Thailand near a tropical forest (Nakhon Phanom province) made by myself Work by Matthias Sebulke AKA Mattes

Unit 1.1.9: Probability

TARGET

Learning Unit 1.1.9 introduces the concept of *probability*, which is due to Jakob Bernoulli, for the quantification of the structure of randomness. The most important aim of this unit is to show that the concept of probability has a unique interpretation in contrast to many textbooks on probability theory and statistics, where several and inconsistent interpretations are given.

CONTENT

Randomness and Determinism

Quantification is a prerequisite for investigating randomness in an unambiguous way, as explained in previous learning unit. However, quantification of randomness assumes that it has been acknowledged as an inherent part of the physical universe necessary for describing uncertainty. Exactly at this point the difficulties start. Since traditional science emerged in Europe more than thousand years ago, it was and is dominated by the dogma of determinism[34], which left no space for randomness. Instead of considering randomness, the search for the truth dominated all scientific efforts. First science aimed at searching and verifying the *determinate truth* of the Holy Scripture and after this period, it changed smoothly to search for the divine *truths* as written in the Book of Nature. Even later, science started to search in imaginary ideal systems being the result of the human dream of possessing the *truth*.

Despite the discoveries of many physical and biological phenomena which cannot be explained by deterministic laws, determinism is still prevailing in contemporary science. The professor of systematic theology S. Paul Schilling ([6] notes (Chance and Order in Science and Theology. *Theology Today* 47, 1-11, 1991):

> "*Obviously, the deterministic postulate dies slowly; it is preserved in name even when redefined to stand for something quite different from its historic meaning. But clearly for increasing numbers of scientists, the traditional notion of a tightly-knit universe with no loose ends is no longer tenable. Even chance is seen to behave in orderly fashion; whether in microscopic particles or everyday complexity it operates to stable, universal laws.*"

If determinism is abandoned from physics in order to detect the rules of randomness, then necessarily the deterministic approach based on the Hamilton formalism must be replaced by something more appropriate based on the notion of probability.

Jakob Bernoulli and the Science of Prediction

When the Swiss mathematician Jakob Bernoulli investigated more than 300 years ago some problems related to gambling posed by Christiaan Huygens, he realized that mankind is in need of a Science of Prediction in order to cope with uncertainty. He named this new science

[34] Determinism goes, at least partly, back to the belief in divine providence.

Stochastics[35] and his masterpiece Ars conjectandi, which was published in 1713 eight years after he had passed away, introduced the fundamental concepts of the new science and the analytical proof that in fact the proposed science can be developed. One of the fundamentals consists of the quantification of randomness by the concept of probability.

The Concept of Probability

Uncertainty refers to the question which of the possible events will occur in future. If the question could be answered, almost no decision problem would exist. However, the question can be answered only after the occurrence, i. e., after the future event has been turned into a fact. However, different events are attracted with different strength by randomness depending on the initial conditions. Thus, the degree of certainty of the occurrence of a given event varies according to the actual force of attraction. Jakob Bernoulli realized intuitively this relation and defined the *probability of an event as the degree of certainty of its occurrence*[36]

This degree is independent of any subjective opinion or knowledge as it is completely determined by the actual conditions and quantifies randomness. The subsequent problem of how to measure a probability was also solved by Jakob Bernoulli. He developed a measurement procedure[37] for experimentally determining the numerical value of the probability of a specified event for given initial conditions.

Let X be a the aspect of interest in the future and d be the given initial conditions. Then the notation $X|\{d\}$ is introduced for the future aspect of interest X subject to the initial conditions d. Let E be a specified event with respect to the aspect X.

Then the probability of the event E is denoted by $P_{X|\{d\}}(E)$. The probability is a degree and, therefore, it may adopt values between 0 and 1, i. e.

$$0 \leq P_{X|\{d\}}(E) \leq 1 \qquad (2)$$

Moreover, if E_1 and E_2 are two mutually exclusive events with respect to X, then the probability must meet the following more or less obvious requirement:

$$P_{X|\{d\}}(E_1 \cup E_2) = P_{X|\{d\}}(E_1) + P_{X|\{d\}}(E_2) \qquad (3)$$

The probability is a characteristic feature of a future event for fixed initial conditions. It reflects the strength of attraction between the initial condition and the event in question.

Probability and Mathematics

From a mathematical point of view, the concept probability is defined as a non-negative (see (2)), normalized (see (2)) and additive (see (3)) set-function on a σ-field, which is generally called a probability measure. Consequently, probability theory as a special branch of mathematics investigates non-negative, normalized and additive set-functions. The axiomatic basis for this branch of mathematics was laid by A. Kolmogorov in 1933 (see [5], p. 2ff). The complete absence of any real-world meaning of the mathematical concept of probability is stressed by Kolmogorov:

[35] Greek: Science of Prediction.

[36] Jakob Bernoulli's definition of probability has been widely looked at as a means for quantifying subjective opinions and not for quantifying objective randomness. This is concluded from the following words in the Ars conjectandi: "In themselves and objectively, all things under the sun, which are, were, or will be, always have the highest certainty." However, Bernoulli continues: "Unless, indeed, whatever will be will occur with certainty, it is not apparent how the praise of the highest Creator's omniscience and omnipotence can prevail. Others may dispute how this certainty of future occurrences may coexist with the contingency and freedom of secondary causes; we do not wish to deal with matters extraneous to our goal." (Translations from Sylla [3], 2006, p.315.) With these words Jakob Bernoulli indicates some doubts without wanting to go into details.

[37] Note that quantification is a necessary condition for measurement.

> *In accordance with the above, in §1 the concept of a field of probabilities is defined as a system of sets which satisfy certain conditions. What the elements of this set represent is of no importance in the purely mathematical development of the theory of probability (cf. the introduction of basic geometric concepts in the Foundation of Geometry by Hilbert, or the definitions of groups, rings and fields in abstract algebra).*

In this context it is worth noting that the axioms of mathematical probability theory do not include random variables in the sense as introduced here. The *random variables* considered in mathematics are defined as measurable functions and, therefore, should not be confused with the identically named quantities in Bernoulli Stochastics.

Probability and Relative Frequency

In many textbooks (e.g. [1], p. 31,32) on probability or statistics, the probability of an event is defined as the limit of the relative frequency of its occurrence in an infinite sequence of independent and identical repetitions of the corresponding experiment. Note that this definition is motivated by the problem of determining the actual value of the probability of a given event and, therefore, is not appropriate for describing the development of evolution, which - obviously - does not depend on any sequence of experiments.

Scientific Ambiguity of the Concept Probability

Because the different nature of past and future is not adequately taken into account in contemporary science, the concept *probability* was not defined clearly leading to many misunderstandings and misuse. George Boole ([4], p. 187) cites Poisson in order to define probability:

> *A distinguished writer*[38] *has thus stated the fundamental definitions of the science:*
>
> > *"The probability of an event is the reason we have to believe that it has taken place, or that it will take place."*
> >
> > *"The measure of probability of an event is the ratio of the number of cases favorable to that event, to the total number of cases favourable or contrary, and all equally possible" (equally likely to happen).*

The first definition opens *officially* the way of *belief* into science. In fact, science is based on belief, however, this fact has always be concealed until probability was introduced as scientific notion of the *degree of belief*. The second definition reduces the concept of probability to a special case, namely to the case of uniform probabilities, which, strictly speaking, can be found only in the field of *gambling*.

References

[1] Bartosyzyński and Niewiadomska-Bugaj, M. (1996): *Probability and Statistical Inference.* Wiley, New York.

[2] Bernoulli, J. (1713): *Ars conjectandi.* Basel.

[3] Bernoulli, J. (2006): *Ars conjectandi together with Letter to a Friend on Sets in Court Tennis.* Translated with an introduction and notes by Edith Dudley Sylla. The John Hopkins University Press, Baltimore.

[4] Boole, G. (1854): *The Laws of Thought.* Macmillan, London.

[38] Poisson, Recherches sur la Probabilité des Jugement.

[5] Kolmogorov, A.N. (1956): *Foundadtions of the Theory of Probability.* 2nd ed., translated by N. Morrison, Chelsea, New York.

[6] Schilling, S.P. (1991): Chance and Order in Science and Theology. *Theology Today*, 47, 365-376.
http://theologytoday.ptsem.edu/jan1991/v47-4-article1.htm (April 2007).

Examples

1. **Probability and Facts**

Facts characterize the completed evolution regarding certain aspects. They exist and, hence, are determinate. As soon as the aspect in question is quantified by a variable D, the corresponding fact can be expressed by a number, say d. One could say that with respect to the aspect quantified by D the event $\{d\}$ has occurred. Thus, the degree of certainty of the occurrence of the event $\{d\}$ is 1.0, as it is a fact.

Assigning any other value than 1.0 as probability to an event which has already occurred is meaningless. Similar, if an event of the completed evolution is considered which has not occurred, i. e., which has not become a fact, then the degree of certainty of its occurrence is 0 and any other value is meaningless.

2. **Tomorrow's Weather**

Weather forecasts are made on numerous observations (see satellite picture at left) and sophisticated calculations by means of the largest and fastest super computers nowadays available. However, although it is obvious that the weather development is subject to randomness, deterministic models based on systems of differential equations are generally used for making the forecasts.

Figure 1: Satellite Picture of Europe.

Regarding tomorrow's weather many aspects could be of interest. For instance:

- Whether or not it will rain during a given time period, say from 10 in the morning to 3 in the afternoon.
- The lowest or highest temperature during a given time period, say from 8 in the morning to 5 in the afternoon.
- The amount of rain during a given time period, say from 6 in the morning to 8 in the afternoon.

The first aspect is quantified by a so-called indicator variable X for the event "rain", which can adopt only two values namely 0 and 1, where 0 stands for the event "no rain" and 1 for the event "rain." There are only three possible events which may occur. These are $\{0\}$, i. e., "no rain", or $\{1\}$, i. e., "rain", or $\{0, 1\}$, i. e., "no rain" or "rain." Clearly, the last event occurs with certainty and hence its probability is 1.0. The probabilities of the two other events depends on the actual weather conditions quantified by the variable D with actual value d. Then we have:

$$P_{X|\{d\}}(\{0\}) = 1 - P_{X|\{d\}}(\{1\}) \tag{4}$$

and the task would be to identify d and to determine the unknown value $P_{X|\{d\}}(\{1\})$ or, equivalently, $P_{X|\{d\}}(\{0\})$.

The other aspects are quantified by selecting an adequate temperature scale and volume scale, respectively. Again the relevant weather aspects have to be quantified by a variable D and its actual value d must be identified yielding the variable $X|\{d\}$ with probability measure $P_{X|\{d\}}$.

3. **Next Year's Gross National Product**

   The state and development of a nation' economy is often measured by its gross national product (GNP) and, therefore, predictions of the GNP play an important role in many decision making processes. However, as long as the probability of such a predicted event is not known, the predictions is more or less meaningless and should not be used for making decisions.

   Quantification of the GNP is done by selecting an appropriate money unit leading to a random variable X. Much more difficult is the identification of the relevant factor D, which constitute the deterministic variable of the starting situation and specifying its actual value d.

   Once the actual random variable $X|\{d\}$ has been obtained, the corresponding probability measure must be determined as a necessary condition for making useful predictions.

Unit 1.1.10: Ignorance

Target

Ignorance about the actual facts is the major source of human uncertainty about the future development. Unit 1.1.10 aims at showing how human ignorance can be quantified in order to handle ignorance and to purposefully reduce it, where reducing ignorance represents a learning process.

Content

Human Ignorance

This unit deals with *human ignorance*, because Bernoulli Stochastics is a human invention aiming at dealing with human uncertainty. Moreover, only little is known about the ignorance of animals and plants.

While randomness is a characteristic property of evolution, ignorance is a characteristic property of the animated nature and, particularly, of mankind. Randomness takes care of order by generating a structured variability with respect to *what will be*. In contrast, *human ignorance* refers to *what is* and generates an unstructured seeming variability.

Ignorance is the opposite of knowledge. Knowledge and, hence, ignorance refer to facts, i. e., to the past, i. e., to the completed evolution. In case of ignorance, the facts on hand quantified by the variable D are not known. Knowledge refers to *what is not* and enables to exclude almost all values for the unknown d except for a bounded set of values.

Variability Generated by Ignorance

Similar as randomness generates variability with respect to the future outcome, ignorance generates variability with respect to the past. However, in contrast to the indeterminate future, the past is fixed and, therefore, the variability generated by ignorance exists only seemingly.

Moreover, if in a thought experiment a probability structure should be assigned to the set of potential states, then the actual state gets the probability value 1 and to all the other states the probability value 0. Such a *deterministic* probability distribution is sometimes referred to as a degenerate probability distribution or a one-point probability distribution.

Because the actual state is unknown, the existing (degenerate) structure is unknown. Preferring some state to other states in such a situation would be tantamount to relying on subjective belief, which would violate the aim of science to overcome human subjectivity. Therefore, no state within the set of potential states must be preferred.

The only way to reduce the seemingly existing variability is to reduce ignorance by means of a learning process.

Learning Means Exploiting Randomness

Ignorance is quantified by the set of potential states which includes the actual one. This set is necessarily bounded and obtained by excluding all those states which cannot be the actual one.

Of course, excluding states assumes that there is some knowledge about the actual situation. However, there is no situation thinkable where it is impossible to state explicitly a finite or bounded set which covers the initial condition, i.e., the actual situation.

Reducing ignorance assumes necessarily a learning process, i. e., a process which is started and which (hopefully) results after completion in the exclusion of some more elements of the set of potential states. At the start of the learning process its outcome is indeterminate. As soon as the outcome of the (random) learning process is available, those elements of the set of potential states can be excluded, which are not consistent with the obtained outcome. Thus, learning is possible only by performing a process with random outcome, which shows that all our knowledge is based on randomness.

The Learning Process

Learning about the past means to use the stochastic relation between past and future for reducing the set of potential states and, thus, ignorance. The relation between past and future indicates those pairs of initial conditions and future outcomes, which are compatible and those which are incompatible. Thus, each possible outcome of the learning process allows to exclude the incompatible states from the set of potential states. If the set of potential states does not include any of the incompatible states, the learning process was in vain, otherwise it was successful. The more states can be excluded the larger was the learning success.

Examples

1. **Ignorance and Belief**

   A professor of theology at the University of Würzburg kept saying that "belief is knowledge which we rely in," however, without ever specifying the meaning of knowledge. Defining "knowledge" by "knowing the truth", it is obvious that mankind will never reach the state of knowledge. Therefore, it makes much more sense to operate with the concept of ignorance for describing the state of human insight. Using ignorance as basic concept leads immediately to the following statement about belief: "Belief is ignorance which is not acknowledged as such." In fact belief is one of the most dangerous forms of ignorance, since many people tend to impose their own belief upon those who believe in something different.

2. **Measurement and Ignorance**

   Determination of facts is usually called "measurement," at least if the facts are quantified. Thus, for reducing ignorance about the actual values of e.g. weights, temperatures, areas, etc. measurement procedures and measurement devices are used. These devices exploit some random process which is more or less closely related to the fact which shall be assessed. If the measurements are repeated several times, then generally different outcomes will be observed. Moreover, a measurement result may be obtained by different facts. In other words, there is no measurement procedure by means of which it is possible to determine exactly the given fact. It is only possible to exclude more or less many of the values which must be considered as potentially possible at the start of the measurement. The set of values which after the measurement still cannot be excluded represent the reached state of ignorance about *what is*, or the state of knowledge about *what is not*.

3. **The Learning Process and Ignorance**

   Clearly, any learning process aims at reducing ignorance. Moreover, any learning process is a random experiment with many possible outcomes. Consider the process of learning a foreign language. It is well known that the learning success depends decisively on the persons. Some learn extremely fast and well, while others fail completely. Thus, the individual capability is one of the main relevant factors which have to be quantified by means of the deterministic variable D. Besides the individual capability, the design of the learning process is another relevant factor. To increase the probability of success the learning experiment should be designed by taking into account the individual capability, the design of the learning process, and many other factors.

Unit 1.1.11: Knowledge

Target

The Learning Unit 1.1.11 aims at preparing for the second module by elaborating the next task to be performed in Bernoulli Stochastics. The next task is closely related to knowledge and ignorance. While science searches for truth, Bernoulli Stochastics just aims at reducing ignorance. The second goal of this unit consists of showing that mankind's knowledge about the *evolution* is extremely limited and that other creatures have possibly a better understanding of evolution than human beings.

Content

Object of Knowledge

Evolution advances continuously and yields a permanently changing universe. Knowledge refers to facts, i. e., to the completed part of evolution and the question arises which facts are relevant, if a future development, i. e., a part of uncompleted evolution is of interest. Because of the universal connectivity, the completed part of evolution determines the extent of future variability on the one hand and the structure of randomness on the other. Thus, due to the connectivity also extent and structure of future variability are facts and, therefore, may be part of human knowledge implying that both should be the objects of knowledge.

Acquiring Knowledge

The continuously increasing completed part of evolution includes the extent and structure of future variability and represents *what is*. The universal connectivity enables man to catch a limited glimpse on the past, but prevents that the completed part of evolution, i. e., (*what is*), will ever be entirely identified. This ever changing *what is* constitutes truth implying that mankind will never be able to acquire truth. As to evolution, even the extremely limited human experience allows to exclude that the states of universe at two different points of evolution are identical and, therefore, even if the universe would be deterministic and the truth would be known an exact prediction about the course of the uncompleted part of evolution would not be possible.

Mankind has no direct access to facts, but accumulates knowledge indirectly by identifying *what is not*, i. e., not by observing existing facts, but by excluding not existing states. For example, there is no measurement device, which yields a true value, i. e., an existing fact. By measuring something it is only possible to exclude those values which are not compatible with the outcome of the measurement process. Thus a meaningful measurement result must be a set (or interval), which hopefully contains the true value.

The human knowledge acquiring process consists of reducing ignorance by excluding states, which are not consistent with a given situation. The result is a set containing necessarily more than one state of universe. The larger the set of potential states, which cannot be excluded, the larger ignorance and the lesser knowledge are.

Thus, knowledge is the compliment of ignorance and just as man cannot acquire complete knowledge, man cannot abandon ignorance completely and, therefore, believing to possess truth constitutes always a severe error. Thomas Jefferson shall have said (www.wisdomquotes.com):

> *Ignorance is preferable to error, and he is less remote from the truth who believes nothing than he who believes what is wrong.*

Improving Knowledge

Improving knowledge is equivalent of reducing ignorance and means to exclude further states from the set of potential states, which so far could not be excluded. Thus, learning means to approaching "truth" without ever reaching it. Pretending of having reached "truth" is equivalently of self-delusion and, therefore, of a failure of the learning effort. Evidently, excluding further states, i. e., improving knowledge, is possible only if the set of potential states contains more than one element. A singleton as set of potential states, which is tantamount to the belief of possessing truth, makes any continuous learning process impossible. The late Librarian of (U.S.) Congress Daniel J. Boorstin has expressed this fact in the following words (www.wisdomquotes.com):

> *The greatest obstacle to discovery is not ignorance - it is the illusion of knowledge.*

Representation of Knowledge

The main problem which has to be solved is to find an appropriate representation of the state of knowledge or equivalently of the state of ignorance. Note that *belief* is not considered as knowledge here, because belief is not the result of an exclusion process, but the result of individual preference leading to a seeming truth, which, if quantified, can be represented by a singleton.

As above stated knowledge refers to facts. These facts also include features of the future development namely extent and structure of variability. What is needed, is a quantified representation of "what is known" or more precisely "what is unknown" about the relevant facts with respect to the question of interest.

Therefore, before one can learn or acquire knowledge the question should be answered how knowledge or ignorance can be represented. Unfortunately, this question has not been answered so far within the realm of traditional science. Ignorance has never been quantified and, consequently, mankind failed to develop methods of purposefully reduce the existing ignorance. The obvious results of this shortfall are wars, terror and catastrophes, which hit the human communities as well as the human individuals.

The Pitfall of Human Learning

Mankind is proud about its capabilities of learning and believes that these capabilities distinguish human beings from the rest of the animal kingdom. George Box[39] presents three different stages of learning and illustrates them by the following figure:

[39] Box, G. (1997): Scientific Method: The Generation of Knowledge and Quality. *Quality Progress*, January 1997, 47-50.

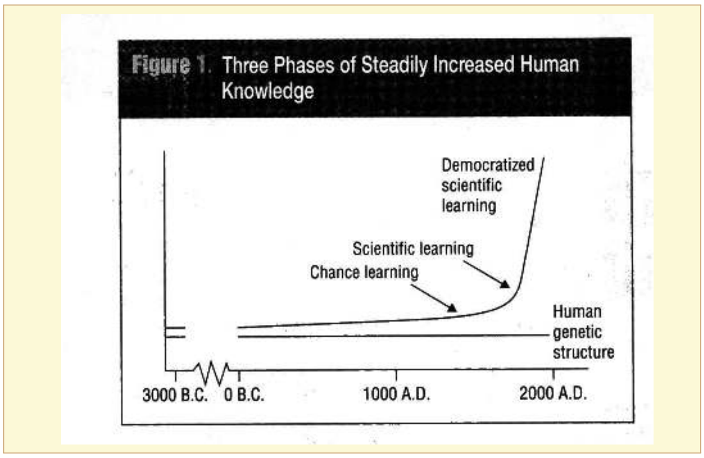


Figure 1: The three phases of human knowledge according to Box.

Box explains his figure as follows:

> *History shows that humans have always learned, but progress once depended on the chance coming together of a informative event and a perceptive observer. Scientific method has accelerated that process in at least four ways:*
>
> - *Providing in a better understanding of the interactive nature of learning.*
> - *Deducting the logical consequences of a group of facts, each individually known but not previously brought together.*
> - *Passively observing and analyzing systems already in operation and data coming from the systems.*
> - *Deliberately staging artificial experiences by experimentation.*

As to the third stage of democratized scientific learning, Box states:

> *A special contribution for which the quality movement is responsible is the democratized and comprehensive diffusion of scientific methods. ... Providing the work force with simple tools (such as control charts, flow-charts, and SPC[40] tools) and empowering their use can ensure that thousands of deductive-inductive brains are actively producing information on how products and processes can be improved.*

As a matter of fact the methods of human learning have not changed since pre-diluvial times and rest still on chance. Michael Polanyi[41] claims that scientific discovery is an art, and he states[42]

> *Since an art cannot be precisely defined, it can be transmitted only be examples of the practice which embodies it. He who would learn from a master by watching him must trust his example. He must recognise as authoritative the art which he wishes to learn and those of whom he would learn it. Unless he presumes that the substance and method of science are fundamentally sound, he will never develop a sense of scientific value and acquire the skill of scientific enquiry. This is the way of acquiring knowledge, which the Christian Church Fathers described as* fides quaerens intellectum, *'to believe in order to know'.*

[40] SPC = Statistical Process Control.

[41] Michael Polanyi was a Fellow of the Royal Society of England and wrote many books on the philosophy of science.

[42] Michael Polanyi: *Science, Faith and Society.* The University of Chicago Press, Chicago and London: 1964, p. 15.

As long as ignorance is not quantified, but just denied and replaced by belief, mankind will not be able to learn how to manage successfully its problems, but contrary continue to blindly tumble in ever increasing problems and ends in situations like that illustrated by the figure below:

Figure[43] 2: The result of not taking appropriate into account the initial condition.

Examples

1. **Driving Time: Facts**

   Assume that W. Tannenbaum is planning to leave Würzburg at 12 a.m. to drive by car to Frankfurt/Airport. The quantity of interest is the driving time.

   The relevant facts which have an impact on the driving time refer to:

   - W. Tannenbaum,
   - W. Tannenbaum's car,
   - the actual traffic situation,
   - the actual highway condition,
   - the selected route, and
   - the actual weather conditions.

   The values adopted by the above outlined facts represent the initial conditions, which relate to the future development. The question how to select the details and how to determine the actual values of the facts will be discussed in the next modules.

   Knowledge refers to the facts per se and their potential values. If a relevant fact is erroneously considered as irrelevant, any value is assumed to be admitted. Note that strictly speaking any fact affects the ongoing evolution. However, the effect of almost all facts is so small that it can be neglected for most purposes. Therefore, we classify a fact as being irrelevant, if its effect on the future development may be neglected without endangering the purpose.

[43]Taken from: *www.tagesschau.de, 23 January 2007.*

2. **Driving Time: Variability**

   Clearly, the driving time is limited by a minimum and a maximum value, which are fixed according to the given set of facts. For instance, if W. Tannenbaum owns a fast car and likes to drive fast, the minimum driving time is smaller compared with a situation, where Tannenbaum owns an old and slow car and drives never fasten than 80 km/h. If the weather conditions are bad with snow and ice the range of the driving time is moved to larger values. If Tannebaum selects the Highway and the traffic conditions are favorable the range is shifted to the smaller values. If the driving day falls on the end of the summer vacations in the Netherlands than the traffic will be dense and the driving time longer.

   As shown, the initial conditions determine the range of variability. The knowledge about the range of variability is obtained by excluding all those values, which are not compatible with the given situation.

3. **Driving Time: Random Structure**

   Similar as the given facts, i. e., the initial conditions, determine the range of variability of the driving time, they also determine the probabilities of the possible events, i. e., the degrees of certainty of their occurrence. For instance having a fast car and liking to drive fast attracts events consisting of small times, while an old and slow car would lead to the exclusion of these events. Using rural roads, bad weather or bad traffic conditions would again lead to the exclusion of "fast" events.

   Again the initial conditions determine the structure of randomness. Knowledge about this structure must be obtained similar as in the case described above, i. e., by excluding structures that are not compatible with the given situation and selecting those which reflect the available knowledge and do not assume non-existing knowledge.

4. **Search for Truth**

   Byrne[44] describes the human search for truth as follows:

   > *For many years, man has been seeking after truth. He seeks it in many different ways, from many different sources, and he has by this time found a considerable variety of ways in which to describe it. He is not even sure, at time, that there is such a thing as truth or, if truth does exist, that it is attainable. Still, man seeks truth. And, being the orderly creature that he is, he wants truth to be orderly as well. Thus he has always shown a marked tendency to make truth after his own image, so that he might present it in a neat and, if possible, little package. This, of course, has not infrequently required considerable ingenuity. For, whatever truth may be in the final analysis, it seldom appears to us to be neat and orderly. Man realizes this, at least in those moments when he is honest with himself, but the realization disturbs rather than pleases him. He simply does not like an untidy world. Thus, the more he learns about the complexity and intricacy of the world in which he lives, the more he seeks to express what he has learned by means of pregnant words, symbols, and formulas. These are , to be sure, often no more than time- and labor-saving devices by means of which he avoids the unpleasant task of pointing out (at some length, it must be added) that the world really is not as simple as all that.*[45]

[44] Byrne, E.F. (1968): *Probability and Opinion. A Study in the Medieval Presuppositions of Post-Medieval Theories of Probability.* Martinus Nijhoff, The Hague.

[45] Byrne, from author's preface, p. XVIII

Module 1.2: Quantification

Content, Aim and Benefits of the Module *Quantification*

Introductory Remarks

The British physicist Sir William Thomson, Lord Kelvin of Largs (1824–1907) once said:

> *"when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind."*

To express something in numbers means to use mathematics as language for communication instead of using some spoken languages. The latter one have the decisive disadvantage that they are ambiguous and, therefore, not well suited for passing over information, as any information which is expressed by means of a natural language must lead almost necessarily to misunderstandings and to non-ending discussions about an appropriate interpretation.

The content of this module is to derive the rules how and tools for describing a part of evolution by means of the mathematical language. These rules are derived by firstly considering the characteristic properties of evolution and mankind, respectively, and only then selecting the appropriate mathematical tools for expressing them. Proceeding like this avoids the difficulties of interpretation which generally arise if one starts with mathematical concepts and only subsequently tries to adapt evolution to the tools. A prominent example for such difficulties is quantum physics where the mathematical derivations still wait for a generally accepted interpretation[46]. In fact, despite of the Copenhagen interpretation, the discussion about the meaning of the mathematical formalism of quantum mechanics has never stopped. Another prominent example of the difficulty of interpretation of mathematical concepts is given by the mathematical concept "probability." In books of probability theory and statistics a multitude of different interpretations are offered, which, by the way, are one of the reasons for the complete confusion with regard to the treatment of uncertainty.

At the beginning of modern science, there was the belief that evolution follows mathematical laws. From this belief it was concluded that evolution cannot be explained based on observations, but only by deriving the underlying mathematical laws. This has led to a science, which is more or less independent of reality.

Quantification starts with reality and means to to translate the observed details into the language mathematics aiming at being able to communicate information about evolution in a unambiguous way.

Passing this module successfully should enable to describe any part of evolution that is of interest by using the language mathematics. It should also make clear that the traditional methods and rules for describing parts of evolution must fail and must lead to misunderstandings and wrong decisions. Therefore, this module represents the core of ´Bernoulli Stochastics and if applied it leads directly to stochastic science.

A successful passing of this module should further deepen the understanding of the universe

[46] The Copenhagen interpretation of quantum mechanics was developed during the 1920s. It is based on Bohr's correspondence principle, Born's statistical interpretation of the wave function, and Bohr's complementarity interpretation of certain atomic phenomena.

and its continuously proceeding evolution. Particularly, it should motivate to abandon causal thinking in favor of thinking in stochastic relations.

Note that stochastic models, i. e., a quantitative images of parts of evolution, which are derived in this module are completely different than those used in physics. However, physical models are in some sense covered by stochastic models and, therefore, may be used as starting points for developing stochastic models.

Abandoning causal thinking would not only change science, but it would also change most of the decision-making procedures in all parts and levels of human society, which are based on assumed cause-effect relations.

Last but not least the rules developed in this module can be regarded as rules which can be used to distinguish ‘science’ from ‘non-science’.

Unit 1.2.1: Features and Variables

TARGET

The goal of the Learning Unit 1.2.1, is to explain the quantification of the most simplest quantities, namely features or characteristics or attributes. They are quantified by variables, i.e., a mathematical quantity that is defined by a symbol and a set of numbers which is called range of variability.

CONTENT

Quantification and Mathematics

Quantification of a situation means to describe it by means of the mathematical language. There are at least three advantages of using mathematics:

- The situation can be described in an unambiguous way, because mathematics is based on numbers.
- The description is consistent, because mathematics is developed according the rules of logic.
- The mathematical tools for describing an analyzing complex situations become available.

These three advantages are decisive for developing the *stochastic science*. If a description of the situation happens to be ambiguous, then all the subsequent efforts would be focussed on overcoming the ambiguity, while solving the underlying problem would be neglected. The second advantage means that the different features entering the description will not contradict one another. In order to illustrate the third advantage, let us consider a four dimensional situation and let us try to describe its properties and feature by means of any natural language. It is simply impossible, because the perception of mankind is basically restricted to three dimensions. Any further complication lies beyond the human perception and, thus, outside the human (natural) languages. By using mathematics instead, one overcomes this difficulty and the fourth dimension can be added in a more or less natural way, without any additional technical complexity.

The first step of quantification, namely, describing a given process by using mathematics as language, consists of identifying the relevant features (features of interest or relevance) and replacing them by variables. This first step of quantification is extremely important, because it constitutes the base of all the subsequent steps.

Different Types of Features Require Different Rules of Quantification.

A feature can adopt different values and quantification means to assign a number to each of the different values. The rules of how to assign the numbers depend on the nature or type of the feature.

- Nominal Feature:

  The values of a nominal feature represent categories, without any order. There is only one rule for quantification: Each selected number must represent exactly one value.

- Ordinal Feature:

  The values of an ordinal feature define a hierarchy or ranking. Thus, quantification must assign to each value a different number in such a way that the hierarchy or ranking is preserved.

- Metric Feature:

  The values of a metric feature establish a metric, e.g. a distance. The quantification must assign to each value a different number and the numbers must be selected in a way that they preserve the given metric. For a metric feature, two different subtypes are often distinguished.

  – Feature without natural point of origin:

    If the considered feature has no natural point of origin, which corresponds to the zero in mathematics, then zero can be assigned to an arbitrarily selected value. However, in this case ratios of two selected numbers have no real-world meaning.

  – Feature with natural point of origin:

    If the considered feature has a natural point of origin, which corresponds to the zero in mathematics, then zero must be assigned to the distinguished value. In this case, ratios of two selected numbers do have a real-world meaning.

Subsequent Steps of Quantification

The first step of quantification consists of replacing all the involved features by the corresponding variables. In the subsequent steps, the relations between the variables have to be described by functions.

Thus, the complete quantified image of a process involves variables and functions, where the former stands for the features and the latter for the relations between the features.

EXAMPLES

1. **Murder on the Orient-Express**

   The time until Poirot's investigations of the murder are terminated, either because the murderer is identified, or Hercule Poirot (see photo on the right) has admitted his failure, is an aspect of the future development. If it is of interest, it is represented by the random variable X defined by below.

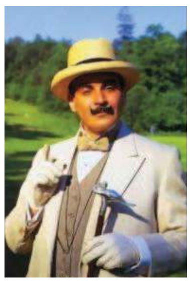


Figure 1: Hercule Poirot

$$X = \text{number of days starting with the day of the murder-event until Poirot's investigation will be terminated}$$

2. **Bowl Filled with Chocolate Balls**

   Consider a child trying to take as many chocolate balls out of the bowl as possible. Clearly, the number of interest is indeterminate, before the process of taking out the balls gets completed. In this case, the following random variable X is identified:

$$X = \text{number of chocolate balls the child will take out of the bowl}$$

3. **Fishing in a Carp pond**

   Assume that the aspect of interest is the weather on a particular fishing day. The question is, whether or not it is going to rain on that day. In this case, the aspect of interest is quantified by a random variable X, which is given by a so-called indicator variable:

$$X = \begin{cases} 0 & \text{if it will not rain} \\ 1 & \text{if it will rain} \end{cases}$$

   A random variable X represented by an indicator variable constitutes the simplest case of an aspect of interest. The entire future is divided into one event and its complementary event.

4. **Solar System**

   Suppose, you are not provided with electric power, so that you could read a book after the sunset. Therefore, you are interested in the time of sunset, when it will get too dark for reading. For this, the random variable X is defined in the following way:

$$X = \text{time in hours and minutes when it will become too dark for reading}$$

Unit 1.2.2: Random Variable

Target

The goal of Learning Unit 1.2.2 is to realize that modeling evolution must necessarily start with identifying the aspects of interest of the future development. Once these aspects have been identified, the first step of quantification can be taken up, where the first step of quantification always refers to the introduction of the necessary variables.

Content

Future Aspect of Interest

Strictly speaking, because of the universal connectivity, evolution encompasses the entire universe and, therefore, is indivisible. However, for mankind almost nothing is of interest except for some very few features of the future development of universe. Clearly, for human individuals the number of features of interest is even much smaller.

From the viewpoint of mankind it is therefore justified to restrict quantification of the future development to the few aspects of interest. Note that identification of these aspects of interest is of utmost importance. They constitute the object of interest and determine almost every further step of quantification process.

Quantification of Future Aspects

For quantification of a future aspect a *variable* is introduced. A variable is one of the elementary concepts of mathematics. It consists of a name (= symbol) and a range of variability which includes all those real numbers[47] which are admitted for the variable.

In Bernoulli Stochastics variables representing future aspects are often symbolized by the capital letter X. The value of the variable X is indeterminate, but restricted to a certain set determined by the actual conditions. This bounded set of possible values which might be adopted in future defines the *range of variability* of X. The symbol together with the range of variability completes the quantification of the future aspect of interest. However, at this first step of quantification, it is sufficient to introduce a symbol, say X, representing the identified aspect of interest.

The real number, which will be adopted by X in future, is selected by randomness in the course of evolution. It does not exist at the start of the considered process. Variables representing future aspects are therefore called *random variables*. Clearly, random variables representing the future development constitute the center of *stochastic science*.

Note, that one should strictly distinguish between a variable and the actual value adopted by the variable. In Bernoulli Stochastics, variables are denoted by capital letters, e.g. X, while the corresponding numbers are represented by corresponding small letters, e.g. x.

[47] There are also complex variables, which may adopt complex numbers. However, complex variable are of no relevance for our purpose.

Examples

1. **Murder on the Orient-Express**

   The time until the investigations of the murder are terminated either because the murderer is identified or Hercule Poirot has admitted his failure is a random variable:

   $$X = \text{number of days starting with the murder until Poirot's investigation will be terminated}$$

2. **The Bowl Filled with Chocolate Balls**

   The number of balls that will be taken by the child is a random variable:

   $$X = \text{number of chocolate balls the child will take out of the bowl}$$

   Figure 1: Delicious chocolate balls.

3. **Fishing in a Carp pond**

   The indicator variable of the event of rain during a future fishing day is a random variable:

   $$X = \begin{cases} 0 & \text{if it will not rain} \\ 1 & \text{if it will rain} \end{cases}$$

4. **Solar System**

   The time of sunset is a random variable since it refers to a future sunset:

   $$X = \text{time in hours and minutes when it will become too dark for reading}$$

Unit 1.2.3: Deterministic Variable

TARGET

After the aspect of future evolution has been quantified by the random variable X, the relevant facts of the completed evolution have to be identified and represented by a variable, say D, which is called deterministic variable.

The aim of modeling uncertainty consists of describing the relation between the past, represented by the deterministic variable D, and the future, represented by the random variable X.

CONTENT

Determinate Past

Evolution advances according to intrinsic laws which heavily depend on the past development, i. e., on the actual achieved state. Thus, foreseeing a future development depends essentially on the amount of knowledge with respect to the actual state.

Perception of mankind is not much refined and, therefore, a rather rough idea about a future development is often sufficient. Consequently, only a rough idea about the actual state turns out to be necessary. A useful description of the actual state does not need to cover all facts in the universe, which would be impossible. Only some few key features must be considered. The influence of all other features lies not within human resolution and is thus irrelevant and can be neglected. Therefore, after having identified and quantified the aspects of interest in the future, the aspects of relevance of the past, i. e., the actual state, have to be identified and quantified.

Quantification of Past Aspects

Just as in the case of introducing random variables, an aspect of the past is quantified by a *variable*, i. e., a symbol together with a range of variability.

The value of any variable representing an aspect of the past is fixed by the completed evolution, i. e., it is determinate. Therefore, variables used for representing past facts are named deterministic variable often denoted by D. The value a deterministic variable D may adopt is fixed, i. e., certain, however, generally unknown.

Past and Future

Note that in stochastic science future is represented by a random variable X and strictly distinguished from past which is represented by a deterministic variable D. The value of X does not exist so far, it will be selected in future (by randomness) and, therefore, it is uncertain. The value of the deterministic variable is fixed by the completed evolution. Therefore, it is not uncertain, but might be unknown.

In any situation which shall be scientifically analyzed (in the sense of stochastic science), one has to identify the random variable X and subsequently the deterministic variable D. The pair of variables (X, D) represent the quantity of interest, X referring to the future and D to the past.

Without having identified the pair (X, D) it is not possible to describe the situation related to the two variables, i. e., that part of evolution which will lead from the past – represented by the deterministic variable D – to the future – represented by the random variable X.

Representation of the Deterministic Variable

The identification of those quantities of the actual state which have to be taken into account when describing a future development constitutes a major problem. Because of the universal connectivity, any fact has an influence on the future development. However, most facts have only a marginal influence on the process at hand and, therefore, can be neglected. Only those quantities have to be taken into account, which have a major impact on the random variable X. However, this raises new questions:

- What distinguishes a major from a minor impact?
- How does an impact on X become manifest?

These questions will be dealt with in subsequent learning units. Note that the representation of the deterministic variable D is by no means unique. Any one-to-one transformation of a deterministic variable D defines a new equivalent deterministic variable. The same statement holds, of course, also to the random variable X, however, in a different sense. The random variable X represents an aspect of interest and, therefore, transformations are restricted to something like scale-transformations. For example if a time is of interest, one can measure it in days, hours or minutes, but it makes no sense to transform it to something of completely different nature. In contrast, the deterministic variable is not of primary interest and there are no limitations for the transformation. Thus, one could search for a transformation which reduces the number of key factors in order to simplify the situation.

At the time being, it is assumed that the key factors for the process in question have been identified and are denoted by the deterministic variable D.

Examples

1. **Murder on the Orient-Express**

   In order to identify the murderer Hercule Poirot decides to have each suspicious person take part in a test for excluding the innocent persons by watching their reactions quantified by the random variable X. Poirot designs the test in a way that innocent persons will react differently than the murderer. The murderer constitutes one of the key factors and is represented by a deterministic variable D. The quantification of D could be done by numbering the involved persons $(1, 2, \ldots, n)$, which would assign a specified, but unknown integer to the murderer:

   $$D = \text{number of the murderer}$$

2. **The Bowl Filled with Chocolate Balls**

   Let the random variable X be the number of chocolate balls the child will take out of the bowl. The size of the child's hand is one of the key factors represented by a deterministic variable:

$$D \quad = \quad \text{size of the child's hand in cm}^2$$

3. **Fishing in a Carp Pond**

   Let the random variable X be the number of carps you will fish during the morning. Then one of the key factors represented by a deterministic variable is the number of carps in the pond.

$$D = \quad \text{number of carps in the pond}$$

4. **Solar System**

   Assume that the time until the next lunar eclipse at a specified location is of interest. Then the actual constellation of moon earth and sun are key factors:

$$D \quad = \quad \text{relative position of moon, earth and sun}$$

Unit 1.2.4: Ignorance Space

TARGET

The major source for man's uncertainty about the future development is human ignorance. In Learning Unit it is shown how human ignorance can be built into the model. This is of utmost importance, because reducing uncertainty means particularly reducing ignorance.

CONTENT

Ignorance and Knowledge

Ignorance refers to the actual value d_0 of the deterministic variable D which is generally unknown. However, it can be excluded that nothing at all is known about the relevant facts, because in such a case one should not try to model the resulting uncertainty.

Thus, by means of the knowledge about *what is not*, it is always possible to exclude almost all real numbers D could adopt in the given situation and to specify a bounded set of potential values for the deterministic variable D denoted $\mathcal{D}$ meeting the following conditions:

- Let d_0 be the actual value, then $d_0 \in \mathcal{D}$.
- If $d \in \mathcal{D}$, then it cannot be excluded that d is the unknown actual value.

The larger $\mathcal{D}$ is, the larger is the ignorance at hand. Therefore, the set $\mathcal{D}$ is called Ignorance Space. Note that the existing knowledge refers to *what is not*, but not to *what is*.

The ignorance space $\mathcal{D}$ is necessarily bounded, i. e., has a finite size, as infinity can always be excluded as an actual value of a fact. The size of the ignorance space denoted by $|\mathcal{D}|$ is defined as follows:

$$|\mathcal{D}| = \begin{cases} \text{number of elements of } \mathcal{D} \text{ for a discrete set } \mathcal{D} \\ \text{Lebesgue measure of } \mathcal{D} \text{ for a continuous set } \mathcal{D} \end{cases}$$

No Ignorance

In the case of complete knowledge or no ignorance at all, the *Ignorance Space* is a singleton containing only the true or actual real number d_0.

Strictly speaking, complete knowledge (no ignorance) would be tantamount of knowing the "truth" and, therefore, it is not possible. However, the deterministic variable D consists of only those *key-factors*, which have a substantial impact on the indeterminate outcome of the quantity of interest. Often, there are only very few key-factors to be taken into account and, therefore, the case of complete knowledge about the key-factors may actually occur.

Ignorance and Belief

In Bayes Statistics a parameter space is considered which is similar to the Ignorance Space. Moreover, a so-called prior distribution over the parameter space is assumed assigning to each element of the parameter space a *degree of belief* which, subsequently, is taken into account.

One of the major aims of (stochastic) science is to overcome subjectivity (opinon and belief) in order to meeting some reliability requirements. Bayes Statistics does not try to overcome subjectivity, but makes it an inherent part of its methods. It follows that applying Bayesian methods is tantamount of not knowing the risks and, therefore, applying Bayesian methods is simply hazardous. In Bernoulli Stochastics each element of the Ignorance Space $\mathcal{D}$ might be – according to the available knowledge – the actual one and, therefore, each element has to be taken into account equally.

Ignorance and Mankind

Ignorance is a characteristic feature of man and one aim of science is to reduce ignorance. However, pretending that the "truth" is known as it is done in traditional science, does not leave room for reducing ignorance. Therefore, describing a given situation in stochastic science starts with stating explicitly the existing ignorance.

EXAMPLES

The following examples refer to deterministic variables considered in the examples of the previous learning unit.

1. **Murder on the Orient-Express**

   Before starting the test Hercule Poirot takes all known facts for excluding as many of the involved persons from the suspicion of being the murderer as possible. The remaining six persons are numbered from 1 to 6 yielding the following ignorance space:

   $$\mathcal{D} \;=\; \{1, 2, 3, 4, 5, 6\}$$

   Each element of the ignorance space represents one of the suspects and with certainty one element is the murderer.

2. **The Bowl Filled with Chocolate Balls**

   The size of the child's hand is determined by estimating its length (8.0 cm ± 1.0cm) and width (4.0 cm ± 1.0 cm). Thus, the ignorance space with respect to D is given by:

   $$\mathcal{D} \;=\; \{d \in \mathbb{R} \,|\, 21.0 \leq d \leq 45.0\}$$

   Note that admitting every real number between 21.0 and 45.0 for d constitutes an approximation, because neither length nor width are continuous quantities.

3. **Fishing in a Carp Pond**

   Assume that an upper bound of the actual carp population is given by the number N of carps released into the pond. A trivial lower bound is, of course, 0. Thus, the following ignorance space is obtained:

   $$\mathcal{D} = \;\{d \in \mathbb{N} \,|\, 0 \leq d \leq N\}$$

Unit 1.2.5: Variability Function

TARGET

Besides ignorance uncertainty is nurtured by the variability exhibited by evolution. This learning unit is devoted to quantify the amount of variability and to show that ignorance about facts and variability of future developments are completely different.

CONTENT

Random Variable and State of Ignorance

The relationship between past and future affects especially the range of variability of the random variable X. Generally, each potential value of $d \in \mathcal{D}$ determines a different range of variability of the random variable X.

Moreover, also the existing ignorance about the relation between past and future has to be considered when regarding the future variability.

Any state of ignorance with respect to the deterministic variable D is represented by a subset $\mathcal{D}_0$ of the ignorance space $\mathcal{D}$. Let $\mathcal{T}_D(\mathcal{D})$ denote a suitable system of subsets of $\mathcal{D}$, which generally includes

- the singletons $\{d\}$ with $d \in \mathcal{D}$ and
- the entire ignorance space $\mathcal{D}$.

Each element $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ generates a random variable denoted by $X|\mathcal{D}_0$, which represents the aspect of future X on the condition that one of the elements of $d \in \mathcal{D}_0$ is the actual one. In fact, $X|\mathcal{D}_0$ stands for the set of random variables $X|\{d\}$ with $d \in \mathcal{D}_0$. The set of all values which may be adopted by these random variables is called *range of variability* of $X|\mathcal{D}_0$.

The Variability Function

The function which assigns to each considered state of ignorance $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ the corresponding range of variability of the random variable $X|\mathcal{D}_0$ is called variability function denoted by $\mathcal{X}$ with

$$\mathcal{X} : \mathcal{T}_D(\mathcal{D}) \to \mathcal{T}_X(\mathbb{R}^s) \tag{5}$$

where $\mathcal{T}_X(\mathbb{R}^s)$ is a suitably selected system of subsets of the set of s-dimensional real numbers and s is the number of components of X.

The selection of the systems of subsets $\mathcal{T}_D$ and $\mathcal{T}_X$ is an important issue and depends on the aims of the investigation and also on mathematical feasibility. In the one-dimensional case, the systems of intervals are of special importance. However, the types of intervals differ with the aims, as will became clear in the learning units about predictions.

For the singletons $\{d\} \in \mathcal{T}_D(\mathcal{D})$ the values $\mathcal{X}(\{d\})$ of the variability function $\mathcal{X}$ denote the set of possible values of the random variable $X|\{d\}$, while for the entire ignorance-space $\mathcal{D} \in \mathcal{T}_D(\mathcal{D})$

the image $\mathcal{X}(\mathcal{D})$ gives the overall range of variability of X for any of the admitted initial conditions.

Obviously, for any $\mathcal{D}_0 \in \mathcal{T}_D(\mathcal{D})$ the following must hold:

$$\mathcal{X}(\mathcal{D}_0) = \bigcup_{d \in \mathcal{D}_0} \mathcal{X}(\{d\}) \tag{6}$$

where $\mathcal{X}(\{d\})$ is the range of variability of the random variable $X|\{d\}$.

Thus, the only 'mathematical' property of the *Variability Function* consists of the additivity (6), which is a natural requirement implied by its meaning. The images of the variability function represent the ranges of variability of the corresponding random variable, and depend on the assumed state of ignorance with respect to the deterministic variable D.

Size of the Range of Variability

The random variable X represents the issue of interest of the future development. Of course, only those values of X should be taken into account, which can actually be observed.

The resolution of any human observation tool is necessarily finite and, therefore, any image of the variability function and, hence, any range of variability of a random variable is finite, too. Therefore, the following relation

$$|\mathcal{X}(\mathcal{D})| < \infty$$

always holds. If instead a continuous set is used, it is a matter of approximation.

Range of Variability and Ignorance Space

The Ignorance Space includes all those initial conditions, which cannot be excluded in a given situation. Only one of them is the actual value of the deterministic variable D and all the others are wrong. Because the actual value is unknown each of the potential values given by the Ignorance Space must be considered likewise, as otherwise one would rely on belief or subjective opinion which would contradict one of the maxims of stochastic science.

The situation with respect to the range of variability of the random variable X is completely different. There is no actual value and any of the possible values can be the outcome of future evolution. However, the initial conditions exert different forces of attraction to different future events. Thus, the possible outcomes should not been considered likewise, but according to their probabilities. Consequently, modeling must be extended, which is done in the next learning unit.

Examples

The following examples refer to the random variables considered in the examples of Learning Unit 1.2.1 and the deterministic variables considered in the Learning Units 1.2.2 and 1.2.3.

1. **Murder on the Orient-Express**

   Because Hercule Poirot must leave as soon as the Orient-Express reaches after three more days Istanbul, an upper bound for the number of days until termination of the investigations is given. Poirot plans the test which hopefully will unmask the murderer the next day and until then no further activities are planned. Because a failure of the test cannot be excluded with certainty, the following variability function is obtained:

   $$\mathcal{X}(\{d\}) \quad = \quad \{1, 2, 3\} \quad \text{for } d \in \mathcal{D}$$

2. **The Bowl Filled with Chocolate Balls**

   From some experiments it is known that a child with hand-size of 21 cm^2 will grab at least two chocolate balls and at most 4 balls. For increasing the lower and upper bound by one ball the size has to be at least 25 cm^2 and to increase the bounds by two balls at least 35 cm^2. An increase of three balls requires a hand size of at least 50 cm^2, which can be excluded for the actual child. Assume further that there are with certainty more than six chocolate balls in the bowl, then the following variability function is obtained:

$$\mathcal{X}(\{d\} = \begin{cases} \{2,3,4\} & \text{for} \quad 21 \leq d < 25 \\ \{3,4,5\} & \text{for} \quad 25 \leq d < 35 \\ \{4,5,6\} & \text{for} \quad 35 \leq d \leq 45 \end{cases}$$

3. **Fishing and Rain**

   Assume that the weather and particularly the question whether or not it will rain is the aspect of interest during the fishing day. In this case the aspect of interest is quantified by a random variable X, which is given by a so-called indicator variable with range of variability:

$$\mathcal{X}(\{d\} = \{0,1\} \quad \text{for } d \in \mathcal{D}$$

   where $x = 0$ means no rain, while $x = 1$ means rain.

4. **Fishing in a Carp Pond**

   You arrive at the carp pond at 9 a.m. and have to leave at latest at 11 a.m., the time for preparing and casting the fishing rod takes at least 5 minutes. The time needed after a fish has bitten takes at least 10 minutes. Moreover, at 10 o'clock you plan to make a break of at least 30 minutes for drinking a beer and having a sandwich. Finally, you have experienced several times that your efforts were in vain and you haven not caught a single fish. Thus:

$$\mathcal{X}(\{d\}) = \{x \in \mathbb{N} \,|\, 0 \leq x \leq 10\} \quad \text{for } d \in \mathcal{D}$$

Unit 1.2.6: Random Structure Function

Target

Learning Unit 1.2.6 is devoted to the quantification of the random structure modelled on the range of variability.

This learning unit shall support the understand of the random structure and the methods how the random structure can be quantified.

Content

Introduction

Randomness is a kind of attraction exercised by the past on the various future events. The strength of attraction refers to the conformance between the actual state and the considered future event. This strength is quantified by the *probability* of the event, which was introduced by Jakob Bernoulli as the degree of certainty of its occurrence. Thus, randomness assigns to each future event a probability. The future events represent the variability and the corresponding probabilities generate a structure on the future developments. This structure makes it possible to order the events and to make statements about probable and improbable future developments.

Quantification of the Random Structure

The structure refers to the future events, i. e., to subsets of the corresponding range of variability and is adequately described by a probability measure, i. e., a set-function which assigns to each subset of the range of variability (representing a possible event) its probability, i. e., a number between 0 and 1.

Random Structure Function

The situation with respect to the random structure is similar as in the case of the variability function. Each state of ignorance generates a set of random variables and each random variable is characterized by the corresponding range of variability and the corresponding probability measure. The range of variability is obtained by means of the *variability function*, while the set of probability measures is obtained by the so-called random structure function denoted $\mathcal{P}$ and given as follows:

$$\mathcal{P} : \mathcal{T}_D(\mathcal{D}) \rightarrow \mathbb{P} \tag{7}$$

where $\mathbb{P}$ is the set of probability measures.

A probability is a feature of a future event and represents the degree of certainty of its occurrence. The random structure of a random variable $X|\{d\}$ over a range of variability $\mathcal{X}(\{d\})$ is given by the totality of probabilities of the possible events, i. e., the subsets of $\mathcal{X}(\{d\})$. A function defined on the system of subsets of $\mathcal{X}(\{d\})$ with co-domain given by a set of probabilities is called probability measure and investigated in mathematical probability theory.

Significance of the Random Structure

The random structure function specifies the relation between any given state and the possible future states. The relation is a stochastic one and hence specified by probabilities. For any given state $\{d\} \subset \mathcal{D}$ and specified future state $\{x\} \subset \mathcal{X}(\{d\})$, the image $\mathcal{P}(\{d\}) = P_{X|\{d\}}$ of the random structure function fixes the degree of certainty $P_{X|\{d\}}(\{x\})$ that $\{x\}$ will actually occur.

Thus, the random structure function specifies the transition probabilities of any potential given state to any of the possible future states. In fact, evolution consists of permanent changes or transitions, which happen in an indeterministic and irreversible way. Therefore, wishing to describe quantitatively evolution, i. e., the sequence of permanent transitions, then this cane be done only by means of transition probabilities and exactly this is done by the random structure function $\mathcal{P}$.

Examples

1. **Quantum Mechanics**

   The first example shall illustrate the fundamental importance of the random structure function for the entire science. The example consists of a citation from one of Nancy Cartwright's books and refers to quantum mechanics and the interpretation of quantum probabilities.

   > *We began with the question, what are quantum probabilities probabilities of? Practitioners of quantum theory have been reluctant to adopt either non-standard logics or non-standard probabilities. They have rejected the first proposal altogether. Quantum probabilities are not probabilities that the system is 'located at r', but rather, as Merzbacher*[48] *says, that it 'will be found at r in a position measurement'. This answer is no more trouble-free than the first. It suppose that an electron passes through neither one slit nor the other when we are not looking. When we do look, there, suddenly, it is, either at the top slit or at the bottom. What is special about looking that causes objects to be places where they would not have been otherwise? This is just another version of the notorious measurement problem, which we first discussed in the last section.*
   >
   > *I find neither of these two conventional answers very satisfactory, and I propose a more radical alternative. I want to eliminate position probabilities for all the classic dynamic quantities. The only real probabilities in quantum mechanics, I maintain, are transition probabilities. In some circumstances a quantum system will make a transition from one state to another. Indeterministically and irreversibly, without the intervention of any external observer, a system can change its state: the quantum number of the new state will be different and a quantum of some conserved quantity – energy or momentum, or angular momentum, or possibly even strangeness – will be emitted or absorbed. When such a situation occurs, the probabilities for these transitions can be computed; it is these probabilities that serve to interpret quantum mechanics.*[49]

   This citation illustrates the unsolvable dilemma of quantum mechanics, which starts with a completely deterministic model (Schrödinger equation) and suddenly by a simple trick arrives at something which is called probability. In order to follow Nancy Carwright's

[48]Eugen Merzbacher, born 1921 in Berlin, American phycisist.
[49]Cartwright, N. (1983), p.178/179.

proposal, it would be necessary to abandon the entire approach and develop something like stochastic quantum mechanics.

The next examples refer to the random variables considered in the examples of Learning Unit 1.2.4 and shall illustrate how to select the random structure function.

2. **Murder on the Orient-Express**

Based on the experiences with Hercule Poirot's capability in clearing up crimes and setting up traps for detecting criminals and his love for striking scenes, one can exclude with certainty all distributions except a monotonic one with maximum for $x = 1$ yielding:

$$P_{X|\{d\}}(\{1\}) > P_{X|\{d\}}(\{2\}) > P_{X|\{d\}}(\{3\})$$

However, specifying the actual values of the above given probabilities is possible only on the basis of observations made in comparable situations by applying a stochastic measurement procedure, which will be dealt with in the next course of Stochastikon Magister.

3. **The Bowl Filled with Chocolate Balls**

In this case one also can exclude compound distributions and among the simple ones the constant and the monotonic types (see Learning Unit 1.3.6), because the two extreme outcomes are realized only in the rare cases of extremely disadvantageous or advantageous circumstances. Thus, arriving at a distribution of uni-modal type, with

$$\begin{array}{ll} P_{X|\{d\}}(\{2\}) < P_{X|\{d\}}(\{3\}) > P_{X|\{d\}}(\{4\}) & \text{for} \quad 21 \leq d < 25 \\ P_{X|\{d\}}(\{3\}) < P_{X|\{d\}}(\{4\}) > P_{X|\{d\}}(\{5\}) & \text{for} \quad 25 \leq d < 35 \\ P_{X|\{d\}}(\{4\}) < P_{X|\{d\}}(\{5\}) > P_{X|\{d\}}(\{6\}) & \text{for} \quad 35 \leq d \leq 45 \end{array}$$

For specifying the above probabilities experiments would be necessary and a stochastic measurement procedure for determining the unknown probability values.

4. **Fishing in a Carp Pond**

You are an experienced angler and know where the fishes use to stay during the morning hours. Moreover, you have selected a bait which the carps in the pond prefer and, finally, the fishing conditions are each morning the same and not subject of abrupt changes. The latter makes it possible to exclude compound distributions, while the former leads to the exclusion of the constant and the monotonic type of probability distribution. Thus, a uni-modal type of distribution must be taken for modeling the random structure:

$$P_{X|\{d\}}(\{0\}) \leq P_{X|\{d\}}(\{1\}) \leq \ldots \leq P_{X|\{d\}}(\{x^*\}) \geq \ldots \geq P_{X|\{d\}}(\{10\})$$

where the probabilities and the mode x^* have to be determined by appropriate stochastic measurement procedures based on the outcome of a measurement process.

Unit 1.2.7: Probability Functions

TARGET

Learning Unit 1.2.7 introduces the major probability functions used for handling problems in calculating probabilities for given events or determining events for given probabilities.

Moreover, some important quantities related to probability distributions are defined.

CONTENT

Introduction

Randomness of an event is quantified by its probability, i. e., the degree of certainty of its occurrence. The structure of randomness refers to the totality of possible future events and, thus, the structure of randomness is quantified by a function which assigns to each of the events the corresponding probability of occurrence. There are a number of equivalent functions derived in probability theory, which may be taken for quantification. Below, these functions and their properties are briefly introduced.

Probability Measure

A probability measure $P \in \mathbb{P}$ over a given (finite) range of variability $\mathcal{X}$ is a set-function which assigns to each subset of $\mathcal{X}$, which represents a future event, a number between 0 and 1, called the probability of the event. Thus, a probability measure quantifies the degree of certainty of occurrence of each element of the set of possible events. Therefore, it must meet the following intuitive properties:

- The probability of the empty set is given by $P(\emptyset) = 0$. The empty set stands for an impossible event and, therefore, a function assigning a positive probability to an impossible event is clearly not appropriate.
- The probability of the entire set $\mathcal{X}$ is given by $P(\mathcal{X}) = 1$. The entire set $\mathcal{X}$ includes everything which may happen in future. Therefore, its occurrence is certain and, thus, its probability must necessarily be 1.
- Let E_1 and E_2 two disjoint events. Then the probability of the union of $E_1 \cup E_2$ is given by $P(E_1 \cup E_2) = P(E_1) + P(E_2)$. Consider, the special case of $E_2 = \mathcal{X} \setminus E_1$, then it is evident that a function P not meeting the additivity property cannot be used for representing the degree of certainty of the involve events.

A probability measure is a set-function and working with set-functions is rather cumbersome. Therefore, instead of the probability measure there are a number of probability functions derived, which are simpler to work with. Assume a given probability measure P on a range of variability $\mathcal{X}$, then each of the following functions are equivalent with P:

The Probability Mass Function

The probability mass function

$$f : \mathbb{R} \rightarrow \{p \,|\, 0 \leq p \leq 1\}$$

is the most simplest of the probability functions. For a given probability measure P it is defined by

$$f(x) = P(\{x\}) \tag{8}$$

Thus the probability mass function assigns to each event represented by a singleton the probability of its occurrence.

The probability mass function f is equivalent to the corresponding probability measure P, because for any event $E \subset \mathcal{X}$ we have

$$P(E) = \sum_{x \in E} f(x)$$

Because $P(\mathcal{X}) = 1$ we have:

$$\sum_{x \in \mathcal{X}} f(x) = 1.0$$

The Probability Distribution Function

The probability distribution function

$$F : \mathbb{R} \to \{p \,|\, 0 \leq p \leq 1\}$$

is defined by

$$F(x) = P(\{y \,|\, y \leq x\}) \tag{9}$$

Thus the probability distribution function assigns to each event of the form $\{y \,|\, y \leq x\}$ the probability of its occurrence.

The probability distribution function F is equivalent to the corresponding probability mass function f and, thus, also to the underlying probability measure P. Let, without loss of generality, the elements of the range of variability be ordered, i. e., $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ with $x_1 < x_2 < \ldots < x_N$ then we have:

$$f(x_i) = \begin{cases} F(x_1) & \text{for} \quad i = 1 \\ F(x_i) - F(x_{i-1}) & \text{for} \quad 1 < i \leq N \end{cases}$$

The Probability Survival Function

The probability survival function

$$\overline{F} : \mathbb{R} \to \{p \,|\, 0 \leq p \leq 1\}$$

is defined by

$$\overline{F}(x) = P(\{y \,|\, y \geq x\}) \tag{10}$$

Thus, the probability survival function assigns to each event of the form $\{y \,|\, y \geq x\}$ the probability of its occurrence.

The probability survival function $\overline{F}$ is equivalent to the corresponding probability mass function f and, thus, also to the underlying probability measure P. Let, without loss of generality, the elements of the range of variability be ordered, i. e., $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ with $x_1 < x_2 < \ldots < x_N$ then we have:

$$f(x_i) = \begin{cases} \overline{F}(x_i) - \overline{F}(x_{i+1}) & \text{for} \quad 1 \leq i < N \\ \overline{F}(x_N) & \text{for} \quad i = N \end{cases}$$

The Upper Probability Quantile Function

The upper probability quantile function

$$Q^{(u)} : \{p \,|\, 0 < p \leq 1\} \rightarrow \mathcal{X}$$

is defined by

$$Q^{(u)}(p) = \min_{x \in \mathcal{X}} \{x \,|\, F(x) \geq p\} \tag{11}$$

Thus, the upper probability quantile function assigns to each probability p the upper bound of the smallest event of the form $\{y \,|\, y \leq x\}$, which has a probability of not less than p. The upper quantile function is in some sense inverse to the distribution function.

The upper probability quantile function $Q^{(u)}$ is equivalent to the corresponding probability distribution function F and, thus, also to the underlying probability measure P, because

$$F(x) = \max_{p} \{p \,|\, Q^{(u)}(p) = x\}$$

The Lower Probability Quantile Function

The lower probability quantile function

$$Q^{(\ell)} : \{p \,|\, 0 \leq p \leq 1\} \rightarrow \mathcal{X}$$

is defined by

$$Q^{(\ell)}(p) = \max_{x \in \mathcal{X}} \{x \,|\, \overline{F}(x) \geq p\} \tag{12}$$

Thus, the lower probability quantile function assigns to each probability p the lower bound of the smallest event of the form $\{y \,|\, y \geq x\}$, which has a probability of not less than p. The lower quantile function is in some sense inverse to the survival function.

The lower probability quantile function $Q^{(\ell)}$ is equivalent to the corresponding probability survival function $\overline{F}$ and, thus, also to the underlying probability measure P, because

$$\overline{F}(x) = \min_{p} \{p \,|\, Q^{(\ell)}(p) = x\}$$

Remarks:

Note that the mass function, the distribution functions and the survival function may be used to calculate the probability of a given event, while the quantile functions may be used to determine an event for a given probability.

The mass function reflects directly the distribution of the probability mass on the corresponding range of variability. In the univariate case, i. e., for a one-dimensional random variable X, there are simple distributions and compound distributions referring to simple or compound processes. Within the set of simple distributions one can distinguish three types. Consider the random variable X, then we have:

- Type 1: Constant probability distribution with the probability mass function adopting the same value for each possible outcome.
- Type 2: Monotonic probability distribution with a monotonic decreasing or monotonic increasing probability mass function.
- Type 3: Uni-modal probability distribution with a probability mass function which first monotonously increases until an inner maximum is adopted and subsequently decreases.

EXAMPLE

1. **Murder on the Orient-Express**

   Ignorance is large about the probability distribution of the random variable $X|\{d\}$ representing the time until Hercule Poirot's investigation will be terminated. It is only known that

   $$P_{X|\{d\}}(\{1\}) > P_{X|\{d\}}(\{2\}) > P_{X|\{d\}}(\{3\})$$

   holds.

   One possible probability mass function is the following:

   $$f_{X|\{d\}}(x) = \begin{cases} 0.900 & \text{for} \quad x = 1 \\ 0.075 & \text{for} \quad x = 2 \\ 0.025 & \text{for} \quad x = 3 \end{cases}$$

   with the corresponding probability distribution function

   $$F_{X|\{d\}}(x) = \begin{cases} 0.900 & \text{for} \quad x = 1 \\ 0.975 & \text{for} \quad x = 2 \\ 1.000 & \text{for} \quad x = 3 \end{cases}$$

   and survival function

   $$\overline{F}_{X|\{d\}}(x) = \begin{cases} 1.000 & \text{for} \quad x = 1 \\ 0.100 & \text{for} \quad x = 2 \\ 0.025 & \text{for} \quad x = 3 \end{cases}$$

   Finally the corresponding upper and lower quantile functions are as follows:

   $$Q^{(u)}_{X|\{d\}}(p) = \begin{cases} 1 & \text{for} \quad 0 \le p \le 0.900 \\ 2 & \text{for} \quad 0.900 < p \le 0.975 \\ 3 & \text{for} \quad 0.975 < p \le 1.00 \end{cases}$$

   $$Q^{(\ell)}_{X|\{d\}}(p) = \begin{cases} 1 & \text{for} \quad 0.100 < p \le 1.000 \\ 2 & \text{for} \quad 0.025 < p \le 0.100 \\ 3 & \text{for} \quad 0.000 \le p \le 0.025 \end{cases}$$

Unit 1.2.8: The Moments of a Probability Distribution

TARGET

This learning unit introduces the moments of a probability distribution and shows that they can be used as a generic representation of the initial conditions. In other words, the moments of a random variable X can be used as convenient deterministic variable D, which represent the holistic approach in stochastics.

CONTENT

Introduction

The initial conditions represented by the actual value d of the deterministic variable D determine uniquely the corresponding probability distribution $P_{X|\{d\}}$, which is used to quantify the random structure of X in the given situation $\{d\}$. Observing D in a given situation assumes that the key factors are known, a requirement which often is not met in practice. Moreover, even if the key factors are known and, thus, can be observed, the relation between the key factors and the probability distribution are generally unknown, which leads to another unpleasant problem. It would therefore be extremely convenient, if it would be possible to transform D into an equivalent representation, which has a known relation to the probability distribution and which can directly be observed. In order to simplify the considerations, it is assumed here that the random variable X has only one component implying that the range of variability $\mathcal{X}(\{d\})$ is a subset of the set of real numbers $\mathbb{R}$.

Relation between D and $P_{X|\{d\}}$

Let D represent all factors having an impact on the future development given by the random variable X. Let d be the actual value of D, then

$$d \Leftrightarrow P_{X|\{d\}}$$

where the range of variability $\mathcal{X}(\{d\})$ is necessarily bounded.

For a bounded range of variability $\mathcal{X}(\{d\})$, there is another deterministic variable, say D_μ, the actual value μ of which determines uniquely the probability distribution $P_{X|\{d\}}$ on $\mathcal{X}(\{d\})$:

$$P_{X|\{d\}} \Leftrightarrow \mu$$

implying that for given $\mathcal{X}(\{d\})$ we have:

$$d \Leftrightarrow \mu$$

The deterministic variable D_μ consists of the so-called moments of $P_{X|\{d\}}$ or of $X|\{d\}$, where the nth moment denoted by $E\left[(X|\{d\})^n\right]$ for $n = 0, 1, \ldots$ is defined by

$$E\left[(X|\{d\})^n\right] = \sum_{x \in \mathcal{X}(\{d\})} x^n f_{X|\{d\}}(x)$$

The probability distribution $P_{X|\{d\}}$ determines uniquely the sequence $\mu = (\mu_0, \mu_1, \mu_2, \ldots)$ and vice versa, where necessarily $\mu_0 = 0$ holds. Therefore, even if the key factors and hence the deterministic variable D are not known and, consequently, cannot be observed, the probability distribution of $X|\{d\}$ can be determined by observing the corresponding moments $E[X|\{d\}]$.

The nth moment of a random variable $X|\{d\}$ gives the barycenter of the set $\{x^n \mid x \in \mathcal{X}(\{d\})\}$ with respect to the probability mass. The barycenter of a set specifies its location and, therefore, the first moment describes the location of $\mathcal{X}(\{d\})$ or of $X|\{d\}$. This is the reason, why the first moment is often called a measure of location for the considered random variable.

The First Moment $E[X]$

The first moment $E[X]$ of a random variable X is often called expectation of X. The name is not really suitable, because μ_1 is in general not an element of the range of variability of X and, therefore, should not at all be expected to occur.

The term *expectation* is rather old, however, it was not used neither in the famous correspondence between Blaise Pascal and Pierre de Fermat nor by Christiaan Huygens in the original Dutch version of his treatise on gambling. However, in the Latin translation De ratiociniis in ludo aleae of Huygens book, which was published by Huygens teacher van Schooten[50] in 1657 as an appendix of Schooten's *Exercitationum Mathematicarum Libri Quinque*, the word *expectatio* is used, which was translated to English by *expectation*. The same word appears in De Moivre's *Doctrine of Chances* (1718) and the later works.

The expression gained further popularity in the context of the St. Petersburg Paradox, which had an infinite expectation and which was discussed in the early 18th century by the most famous contemporary scientists. In a letter of Gabriel Cramer[51] dated 21st May 1728 the expression "l'espérance mathématique" was used probably for the first time, which is the French equivalent for "mathematical expectation."

The *expectation* or *expected value* were used in the field of gambling as a fair game was defined by equal expectations for each player. All the first books on probability dealt with gambling and fair games and, therefore, with the concept of expectation. In the sequel the expectation gained an inappropriate importance in probability theory and its application and very often interest is focussed exclusively on the expectation ignoring the fact that an isolated value of the first moment has almost no meaning at all.

In the following those moments of a random variable, which have the largest impact on the probability distribution, are identified.

The Key-Moments

The aim is to determine the probability distribution by means of the corresponding moments. Therefore, the question about the significance of the different moments arises.

In the development of mathematical probability theory as well as its application the first moment $E[X|\{d\}]$ and the second central moment $V[X|\{d\}]$ played an important role. The central moments of $X|\{d\}$ are defined as the moments of the random variable $X|\{d\} - \mu_1$, where $\mu_1(d)$ denotes the value of $E[X|\{d\}]$. Because of

$$V[X|\{d\}] = E[X^2|\{d\}] - (E[X|\{d\}])^2$$

the pair of variables $(E[X|\{d\}], E[X^2|\{d\}])$ and $(E[X|\{d\}], V[X|\{d\}])$ are equivalent, as the values of the one determine uniquely the values of the other. The following considerations show that the emphasis on the two first moments is well founded.

[50] Frans van Schooten, 1615 – 1660, Dutch mathematician.
[51] Gabriel Cramer, (1704-1752) Swiss mathematician.

Range of Variability

If the moments of a random variable are numbered beginning with $n = 0$, then the value of $\mu_0(d)$ stands for the range of variability $\mathcal{X}(\{d\})$, which must be known necessarily for defining or calculating all the other moments. Thus, we conclude that the range of variability $\mathcal{X}(\{d\})$ has highest significance and, therefore, belongs to the key-moments.

First Moment

For given range of variability $\mathcal{X}(\{d\})$ of $X|\{d\}$ the theoretical values of the first moment are restricted to the interior of $\mathcal{X}(\{d\})$. Any value of the first moment on the boundary of $\mathcal{X}(\{d\})$ would lead to a one-point distribution violating $\mathcal{X}(\{d\})$.

Because the restrictions imposed on the first moment by the range of variability are rather mild, the first moment is generally also one of the key-moments for the probability distribution.

Second Moment

The question arises about the restrictions imposed by the range of variability and the first moment on the second moment. Let

$$\begin{aligned} a(d) &= \min \mathcal{X}(\{d\}) \\ b(d) &= \max \mathcal{X}(\{d\}) \\ \mu_1(d) &= \sum_{x\in\mathcal{X}(\{d\})} x f_{X|\{d\}}(x) \\ \mu_2(d) &= \sum_{x\in\mathcal{X}(\{d\})} x^2 f_{X|\{d\}}(x) \end{aligned}$$

Then it is easy to show that the following relations holds:

$$(\mu_1(d))^2 < \mu_2(d) < (a(d) + b(d))\mu_1(d) - a(d)b(d) \tag{13}$$

Thus the restriction imposed on the second moment by the range of variability and the first moment are considerable. In Figure 1 the restrictions are illustrated by an example. Only values between the two curves are admitted for the second moment $E[X^2|\{d\}]$.

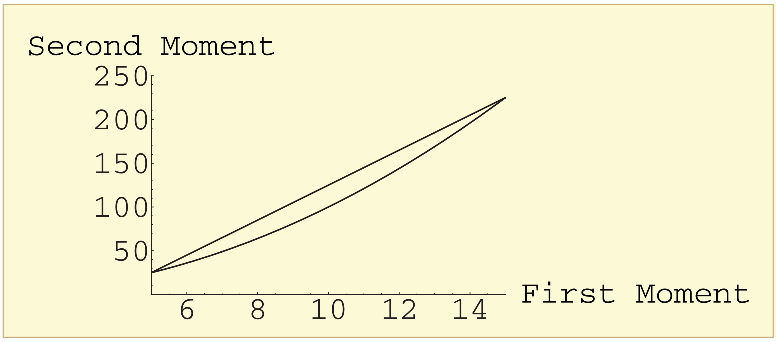


Figure 1: Relations between the first moment and the second moment for the range of variability given by $\mathcal{X}(\{d\}) = \{x \,|\, 5 \le x \le 15\}$.

The restrictions imposed by the range of variability and the first moment on the variance are easily obtained:

$$0 < \sigma^2(d) < -\left(\mu_1 - \frac{a+b}{2}\right)^2 + \left(\frac{a-b}{2}\right)^2 \tag{14}$$

In Figure 2 the relation between the range of variability, the first moment and the variance is illustrated. Accordingly, the variance can adopt only values below the parabola in Figure 2.

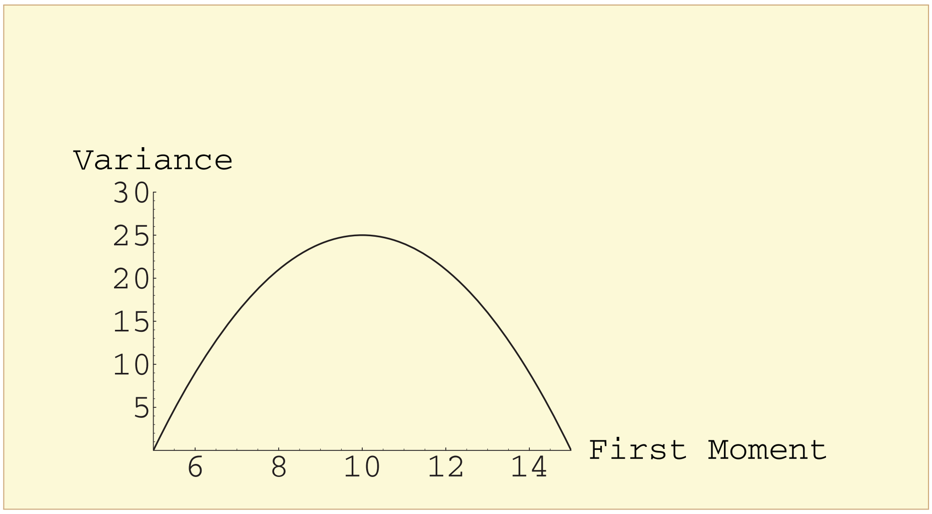


Figure 2: Relations between the first moment and the variance for the range of variability given by $\mathcal{X}(\{d\}) = \{x \,|\, 5 \leq x \leq 15\}$.

The variance has maximum freedom for $\mu_1 = \frac{a+b}{2}$ and gets more and more restricted the more μ_1 approaches one of the boundaries of the range of variability. This behavior of $V[X|\{d\}]$ is intuitively clear. The more μ_1 approaches one of the boundaries, the more is the probability mass concentrated in a neighborhood of the corresponding boundary which limits the value of the second moment or the variance.

In many cases also the second moment or equivalently the variance appears to be a key-moment the value of which must be known for deriving an appropriate probability distribution.

Moment of Higher Order

The restrictions imposed on the moments of order higher than two become very quickly extremely severe, which means that the values of the moments of smaller order determine to a large extent the values of the moments of higher order. Consequently, in many cases the moments of order higher than two are not any more key-moments and they may therefore be neglected.

Examples

The following examples refer to the random variables considered in the examples of the previous learning units.

1. **Murder on the Orient-Express**

   The aspect of interest and, hence, the random variable X is the time measured in days until the end of Hercule Poirot's investigation. The relevant facts or key factors with impact on the future outcome of the test include among many other facts the murderer. They determine the probability distribution over the range of variability. Thus, by changing from the situation-related deterministic variable D to a distribution-related deterministic variable D_p, the following pair of variables with corresponding Bernoulli Space is obtained:

   - Random variable X with

     $$X = \text{number of days until the end of Hercule Poirot's investigation}$$

   - Deterministic variable D_p with

     $$D_p = (D_p^{(1)}, D_p^{(2)})$$

and

$$D_p^{(1)} = \text{probability of the event } \{1\}$$
$$D_p^{(2)} = \text{probability of the event } \{2\}$$

2. **The Bowl Filled with Chocolate Balls**

Similar as in the case of the murder on the Orient-Express, the situation-related deterministic variable including the hand-size, a distribution related deterministic variable D_p is selected. Thus, the following pair of variables with corresponding Bernoulli Space is obtained:

- Random Variable X with

$$X = \text{number of chocolate balls, which will be drawn from the bowl}$$

- Deterministic Variable D_p with

$$D_p = (D_p^{(3)}, D_p^{(4)}, D_p^{(5)})$$

and

$$D_p^{(3)} = \text{probability of the event } \{3\}$$
$$D_p^{(4)} = \text{probability of the event } \{4\}$$
$$D_p^{(5)} = \text{probability of the event } \{5\}$$

Unit 1.2.9: The Stochastic Model

Target

Learning Unit 1.2.9 combines the different quantified pieces and builds up a completely quantified model of uncertainty for a process of interest in a given situation.

The aim is to illustrate the differences between a stochastic model and models which are used in mathematics and traditional science and to show that the conventional models are incomplete and, therefore, should not be used.

Content

Introduction

A description of the relation between past and future that can be used for making reliable predictions must connect the future aspect of interest X and the relevant past facts D. Thus, it refers to a pair of variables (X, D). It must necessarily

- cover human ignorance as a characteristic feature of man, and
- include randomness as the characteristic feature of evolution.

The former is simply described by the amount of ignorance with respect to D, while the latter is represented by the range of variability of X and the corresponding probability structure.

The entire model is called Bernoulli Space denoted by $\mathbb{B}_{X,D}$ and given by

$$\mathbb{B}_{X,D} = (\mathcal{D}, \mathcal{X}, \mathcal{P}) \tag{15}$$

where

- the set $\mathcal{D}$ is called the ignorance space and represents the amount of ignorance,
- the variability function $\mathcal{X}$ assigns to each level of ignorance the amount of future variability, which has to be taken into account,
- the random structure function $\mathcal{P}$ assigns to each level of ignorance a probability distribution.

Note that each of the three components of the Bernoulli Space is necessary for describing uncertainty about the future development of a real-world process. Neglecting ignorance would be tantamount to pretending to knowing the “truth” or to be like God. Not considering the variability and the random structure of the future development would be tantamount to neglecting age-old experiences by believing in determinism, although no deterministic process has ever been observed.

Any mathematical model, which does not include the three components of a Bernoulli Space, is necessarily incomplete, violates reality and is based on belief.

Bernoulli Space versus Probability Space

In 1933 the Russian mathematician A. Kolmogorov succeeded in axiomatizing probability theory and in fact Kolmogorov's axioms constituted a huge step for mathematics as it paved the way for a new, genuine branch of mathematics dealing with measurable functions based on a probability space.

However, for science Kolmogorov's axioms meant a drawback! Kolmogorov's probability space takes into account the needs of mathematics, but neglects totally the needs of science. Kolmogorov's probability space is based on an abstract set – the set Ω of *elementary events* – which has no direct counterpart in reality and which cannot be specified even in the simplest realistic case. The next component of a probability space is a σ-field $\mathcal{S}$ over Ω, which just as Ω cannot be specified for a realistic process. The σ-field meets many mathematical requirements, but from a more realistic point of view, it contains almost only irrelevant elements. Finally, there is a normed, total-additive set-function $P : \mathcal{S} \to \{p \,|\, 0 \leq p \leq 1\}$, which as Kolmogorov states, may have infinite different real-world interpretations.

The probability space does not allow for including ignorance about the relevant facts about the future of interest. Thus, a probability space is not appropriate for describing mathematically what is known about the relation between past and future aiming at controlling human uncertainty about the future development of a part of evolution.

Bernoulli Space versus Mathematical Models in Physics

Physics is basically built up on deterministic models pretending to represent "truth." Similar as the Bernoulli Space, physical models refer to a pair of variables (X, D), where X represents the aspect of interest in the future, and D the initial conditions, i. e., the relevant facts of past. However, physical models assume complete knowledge and consequently the ignorance space $\mathcal{D}$ shrinks to a singleton $\{d\}$. The second assumptions refers to a deterministic evolution which does not allow for any variability of $X|\{d\}$. Thus, the range of variability $\mathcal{X}(\{d\})$ of $X|\{d\}$ also shrinks to a singleton $\{x\}$, while the probability distribution $P_{X|\{d\}}$ of $X|\{d\}$ degenerates to a so-called one-point probability distribution.

To sum up, from a stochastic point of view, physical models constitute more or less trivial and degenerate descriptions of real world completely based on belief and not consistent with any experience or observation.

Bernoulli Space and Science

Mankind has started many attempts for reducing uncertainty about evolution. The most important one, modern science was started during the 16th century. It is believed to be the most most successful one, as it led to the development of modern technology and modern civilization. However, with respect to reducing uncertainty about evolution modern science represents a total failure. Except for some tiny areas in physical sciences no advantages in predicting evolution have been made. Reliable predictions should lead to preventing or solving of problems mankind is or will be faced with. Looking at the history of mankind one has to realize that number and weights of problems have not been reduced. Contrary, local problems have spread out to become global ones and future seems to be more uncertain than ever.

The main reason for the failure of modern science in developing methods for controlling successfully the future development is doubtlessly the neglect of universal connectivity and to focus on isolated systems and cause/effect chains.

Modeling evolution by cause effect chains is as wrong as the attempt to model it according to the Christian Bible. Both approaches must necessarily lead to a distorted way of thinking.

The Bernoulli Space, which includes cause effect chains as degenerated cases, opens the way of accepting man's ignorance and modeling the mysterious and at the same time controlled diversity of evolution.

Bernoulli Space and Stochastic Procedures

The Bernoulli Space enables to compile a list of necessary stochastic procedures:

- Verification

  Procedures are needed for verifying that a Bernoulli Space does not include wrong assumptions.

- Utilization

  A Bernoulli Space is developed aiming at making reliable and accurate predictions. Therefore, prediction procedures are needed.

- Improving

  The predictions can be made more precise, if the actual situation is better known, which immediately leads to an improvement of the stochastic model, i. e., the Bernoulli Space. Thus procedures are needed for

  1. reducing the ignorance space $\mathcal{D}$,
  2. improving the variability function $\mathcal{X}$, and
  3. improving the random structure function $\mathcal{P}$.

As a matter of fact, at the time being only a small part of the desirable procedures are available.

Bernoulli Space and Mathematical Independence

The mathematical concept of *stochastic independence* plays a central role in probability theory and mathematical statistics. Hence, the question arises about the interpretation of the mathematically defined stochastic independence.

In mathematics two elements, say E_1 and E_2 of a given σ-field $\mathcal{S}$ are called *independent* if they meet the condition:

$$P(E_1 \cap E_2) = P(E_1)P(E_2) \tag{16}$$

In order to find a meaningful interpretation of (16), we have to introduce a random variable X, a deterministic variable D and, finally, a Bernoulli Space $\mathbb{B}_{X,D}$ which connects X and D. Then E_1 and E_2 are two subsets of $\mathcal{X}(\mathcal{D})$ and may be interpreted as two possible future events with respect to the random variable X on condition $\mathcal{D}$.

If E_1 and E_2 are disjoint, then only one of them can happen in future. If they are not disjoint both events may happen. If $E_1 = E_2$ then they are equal and with certainty they will happen both or not happen both. Both events E_1 and E_2 refer to the same process specified by X. Comparing the events from a realistic point of view can lead to the result that they are disjoint, that they are similar, or that they are equal, that one is more disliked than the other, that one is more liked than the other, etc., but there is no meaning of stating that E_1 is independent of E_2.

Thus, we conclude that the concept of mathematical independence of events as introduced in probability theory has no meaningful realistic interpretation and must be abandoned within the stochastic model.

Bernoulli Space and Stochastic Dependence

A Bernoulli Space describes mathematically the stochastic relations between past and future and, therefore, gets by without a concept of independence of events or random variables. The relations described by the Bernoulli space are generated by the key-factors which are represented by the deterministic variable D.

Assume that two random variables X_1 and X_2 are given, together with the corresponding deterministic variables D_1 and D_2 and Bernoulli Spaces

$$\mathbb{B}_{X_1,D_1} = (\mathcal{D}_1, \mathcal{X}_1, \mathcal{P}_1)$$
$$\mathbb{B}_{X_2,D_2} = (\mathcal{D}_1, \mathcal{X}_1, \mathcal{P}_1)$$

Next consider the bi-variate random variable $X = (X_1, X_2)$ with deterministic variable D. Let $D_1 = (D_{1,1}, \ldots, D_{1,n_1})$ and $D_2 = (D_{2,1}, \ldots, D_{2,n_2})$. If there are no pairs of components $D_{1,i}$ and $D_{2,j}$ referring to the same fact, then

$$D = (D_{1,1}, \ldots, D_{1,n_1}, D_{2,1}, \ldots, D_{2,n_2})$$

and the corresponding ignorance space is obtained as direct product of the single ignorance spaces:

$$\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$$

As the initial condition with respect to X_1 does not include any key-factor with respect to X_2 and vice versa, the variability function $\mathcal{X}$ is obtained as follows:

$$\mathcal{X}(\{(d_1, d_2)\}) = \mathcal{X}_1(\{d_1\}) \times \mathcal{X}_2(\{d_2\})$$

The joint random structure $\mathcal{P}$ function is obtained analogously:

$$\mathcal{P}\Big(\{(d_1, d_2)\}\Big) = P_{X|\{(d_1,d_2)\}} = P_{X_1|\{d_1\}} \otimes P_{X_2|\{d_2\}}$$

which coincides with the mathematical concept of independent random variables.

To sum up, two random variables X_1 and X_2 are called stochastic independent, if the corresponding deterministic variables have no joint components. In this case the joint ignorance space $\mathcal{D}$ is the direct set product of the ignorance spaces $\mathcal{D}_1$ and $\mathcal{D}_2$, the images of the joint variability function $\mathcal{X}$ are the direct set products of the corresponding images of the variability functions $\mathcal{X}_1$ and $\mathcal{X}_2$ and, finally the images of the joint random structure function $\mathcal{P}$ are the sets of product measures of the corresponding probability measures.

Remark:
Two random variables X_1 and X_2 refer to two aspects of the future development. The values of both variables are not existing at the time being and, therefore, neither of them can influence the other. Thus, stochastic dependence or independence between two random variables refers to the relations of the common past on the development of the two processes represented by X_1 and X_2. If the corresponding key-factors represented by the deterministic variables D_1 and D_2 have joint components, then the always existing dependence between the two considered developments cannot be neglected. For example if the two first moments depend on the same factors in similar manner, then an increase of the value of the first moment of one random variable is accompanied by an increase of the value of the first moment of the second variable.

Examples

The following examples refer to the random variables considered in the examples of the previous learning units.

1. **Murder on the Orient-Express**

   The aspect of interest and, hence, the random variable X is the time measured in days until the end of Hercule Poirot's investigation. The relevant facts or key factors with impact on the future outcome of the test include among many other facts the murderer. They determine the probability distribution over the range of variability. Thus, by changing from the situation-related deterministic variable D to a distribution-related deterministic variable D_p, the following pair of variables with corresponding Bernoulli Space is obtained:

   - Random Variable X with

     $$X = \text{number of days until the end of Hercule Poirot's investigation}$$

   - Deterministic Variable D_p with

     $$D_p = (D_p^{(1)}, D_p^{(2)})$$

     and

     $$\begin{aligned} D_p^{(1)} &= \text{probability of the event } \{1\} \\ D_p^{(2)} &= \text{probability of the event } \{2\} \end{aligned}$$

   - Bernoulli Space $\mathbb{B}_{X,D_p} = (\mathcal{D}_p, \mathcal{X}, \mathcal{P})$ with

     $$\mathcal{D}_p = \Big\{(p_1, p_2) \,|\, 0 \le 1 - (p_1 + p_2) \le p_2 < p_1 \le 1\Big\}$$

     $$\mathcal{X}(\{p_1, p_2\}) = \begin{cases} \{1\} & \text{for} \quad p_1 = 1 \\ \{1,2\} & \text{for} \quad p_1 < p_1 + p_2 = 1 \\ \{1,2,3\} & \text{for} \quad p_3 > 0 \end{cases}$$

     $\mathcal{P}(\{(p_1, p_2)\}) = P_{X|\{(p_1,p_2)\}}$ with probability mass function

     $$f_{X|\{(p_1,p_2)\}}(x) = \begin{cases} p_1 & \text{for} \quad x = 1 \\ p_2 & \text{for} \quad x = 2 \\ 1 - (p_1 + p_2) & \text{for} \quad x = 3 \end{cases}$$

2. **The Bowl Filled with Chocolate Balls**

   Similar as in the case of the Murder on the Orient-Express, the situation-related deterministic variable including the hand-size, a distribution related deterministic variable D_p is selected. Thus, the following pair of variables with corresponding Bernoulli Space is obtained:

   - Random Variable X with

     $$X = \text{number of chocolate balls, which will be drawn from the bowl}$$

   - Deterministic Variable D_p with

     $$D_p = (D_p^{(3)}, D_p^{(4)}, D_p^{(5)})$$

and

$$\begin{aligned} D_p^{(3)} &= \text{probability of the event } \{3\} \\ D_p^{(4)} &= \text{probability of the event } \{4\} \\ D_p^{(5)} &= \text{probability of the event } \{5\} \end{aligned}$$

- Bernoulli Space $\mathbb{B}_{X,D_p} = (\mathcal{D}_p, \mathcal{X}, \mathcal{P})$ with

$$\begin{aligned} \mathcal{D}_p = \Big\{ &(p_3, p_4, p_5) \,|\, 0 < 1 - (p_3 + p_4) < p_3 > p_4 > p_5 = 0 \\ &\vee\ 0 < p_3 < p_4 > p_5 = 1 - (p_3 + p_4) > 0 \\ &\vee\ 0 = p_3 < p_4 < p_5 > 1 - (p_4 + p_5) > 0 \Big\} \end{aligned}$$

$$\mathcal{X}(\{p_3, p_4, p_5\}) = \begin{cases} \{2,3,4\} & \text{for } p_5 = 0 \\ \{3,4,5\} & \text{for } p_3 p_5 > 0 \\ \{4,5,6\} & \text{for } p_3 = 0 \end{cases}$$

$\mathcal{P}(\{(p_3, p_4, p_5)\}) = P_{X|\{(p_3,p_4,p_5)\}}$ with probability mass functions:

$$p_5 = 0: \quad f_{X|\{(p_3,p_4,p_5)\}}(x) = \begin{cases} 1 - (p_3 + p_4) & \text{for } x = 2 \\ p_3 & \text{for } x = 3 \\ p_4 & \text{for } x = 4 \\ 0 & \text{for } x > 4 \end{cases}$$

$$p_3 p_5 > 0: \quad f_{X|\{(p_3,p_4,p_5)\}}(x) = \begin{cases} 0 & \text{for } x = 2 \\ p_3 & \text{for } x = 3 \\ p_4 & \text{for } x = 4 \\ p_5 & \text{for } x = 5 \\ 0 & \text{for } x = 6 \end{cases}$$

$$p_3 = 0: \quad f_{X|\{(p_3,p_4,p_5)\}}(x) = \begin{cases} 0 & \text{for } x = 2 \\ p_3 & \text{for } x = 3 \\ p_4 & \text{for } x = 4 \\ p_5 & \text{for } x = 5 \\ 1 - (p_4 + p_5) & \text{for } x = 6 \end{cases}$$

Remark:

The above two Bernoulli Spaces are characterized by rather large ignorance spaces. Consequently the uncertainty about the future development is too big for making useful predictions possible. In such a case it is necessary to improve the Bernoulli Space by means of learning experiments.

3. **Future Height and Weight of a Person**

In order to illustrate the concept of stochastic independence, take a given child of age $t = 10$ years and consider the following two-dimensional random variable $X = (X_1, X_2)$ with:

$$\begin{aligned} X_1 &= \text{weight of the child at the age of } t = 20 \text{ years} \\ X_2 &= \text{height of the child at the age of } t = 20 \text{ years} \end{aligned}$$

As deterministic variable D certain key-moments $E[X^j X^k]$, for given values of j, k are selected, which determine the joint probability distribution the two-dimensional random variable. The weight of the person depends on its height and therefore the first moment of X_2 depends on the first moment of X_1 and hence the joint distribution function is not the product of the marginal distribution functions.

Unit 1.2.10: Stochastic Science

Target

This learning unit aims at recapitulating the main differences between contemporary science and stochastic science. The learner should understand the necessity for a change to stochastic science in view of efficiency and usefulness of science.

Content

Introduction

The Bernoulli Space is the basis of the development of Stochastic Science. If the goal of science is defined as being the reduction of human uncertainty about future development by making reliable predictions, then the different varieties of science, which have been developed so far, as scholastic science or modern science do not deserve the name "science," because they are based on subjective opinions of authorities and it is therefore impossible to specify the reliability of a prediction.

The scientific theories developed in all branches of contemporary science represent passing fads, which have not been refuted so far, because they do not contain the ever existing ignorance and reflect not sufficiently well the inherently observed variability of evolution.

The escapist nature of contemporary science becomes also manifest by the fact that there are two different sciences necessary: fundamental or pure science and engineering or applied science. Fundamental science produces theories about an imaginary (ideal) world which does not exist. The most acknowledged method in fundamental science is the "flash of genius," while the most common method in engineering science is to proceed by "trial and error." Both methods are extremely inefficient and will hardly lead to a deeper understanding of evolution, although they brought forward an advanced technology. However, in view of the environmental and societal damages caused by these advantages, it is questionable that they represent a real progress for mankind.

There is another aspect concerning fundamental and engineering science. Karl Wulff[52] notes that not technology, but natural philosophy constitutes the root of modern science. Technical progress and the advances of scientific theories proceeded more or less independently in parallel.

Unfortunately, many scientist believe that strict rules or specifications for the advancement of science would ruin science. For example the British professor of physical chemistry Michael Polanyi notes[53]:

> *Admittedly, there are rules which give valuable guidance to scientific discovery, but they are merely rules of art. The application of rules must always rely ultimately on acts not determined by rule. Such acts may be fairly obvious, in which case the rule is said to be precise. But to produce on object by following a precise prescription is a process of manufacture and not a creation of a work of art.*

[52] Karl Wulff (2006): *Naturwissenschaften im Vergleich.* Verlag Harry Deutsch, Frankfurt am Mein, p. 347.
[53] Michael Polanyi: Science, Faith and Society. The University of Chicago Press, Chicago, 1964, p. 14.

Emergence of Stochastic Science

Already a the end of the 17th century Jakob Bernoulli had proposed the development of stochastic science. However, the 17th century is distinguished by a series of ingenious theories about evolution, which looked so much more attractive than Jakob Bernoulli's proposal and since then almost nothing has changed. Mankind still believes in authorities and their brain waves. There are only two evident changes between the 17th century and the 21st century. The number of flashes of genius has decreased, but the expenditures for having one have been increased enormously. As one consequence of this development, the value of scientific research is assessed almost exclusively by the amount of expenditures invested, while the results become more and more insignificant.

Stopping this development is tantamount to completely reformulate aims and rules for science and exactly this is done in Stochastic Science. Science should not deal with imaginary many-worlds' perceptions or imaginary theories on the one hand and it should not be based on trial and error methods, because either of them is too expensive and whether or not they will end in a useful result is also uncertain. Scientific research should proceed in a more purposeful way by defining the aim and starting with deriving the corresponding Bernoulli Space. As soon as the Bernoulli Space is available, the appropriate methods for reaching the aim become more or less evident. As a matter of fact, stochastic science is based on precise rules and, therefore, according to Michael Polanyi not an art but a process of manufacture, and exactly as in manufacturing the products of stochastic science must meet specifications, which shall assure the quality of the products.

Appraisal of Contemporary Science

Clearly, the results of contemporary science are valuable also in the framework of stochastic science, as they reflect certain aspects of the relation between the initial conditions and future outcome of interest. However, a universal appraisal of the available results can not be made. The results can only be evaluated on an individual basis. In any case the results have to be newly interpreted and integrated into the corresponding Bernoulli Spaces.

One obvious difference between contemporary science and stochastic science is that the former is based to a large extent on unrealistic cause-effect relations and the latter on realistic stochastic relations. Explaining a situation by a cause-effect relation often means to walk around wearing blinkers, while stochastic science avoids this shortcoming by always considering the entire situation including ignorance and randomness in order to come to a safe decision.

Another main difference between stochastic science and contemporary science is that stochastic science is open for any process of interest within the inanimate or the animate nature. There is nothing which cannot be described "scientifically" allowing reliable predictions and reducing ignorance. Therefore, stochastic science can be used for investigating scientifically any part of evolution using the same rules and the same methods.

In contrast contemporary science is divided into exact or mathematical sciences and hermeneutical sciences, i. e., in two parts based on completely different methodologies. The exact sciences are restricted at least strictly speaking to those parts of evolution, in which randomness exhibits a relatively small variability. The "exact" methods fail in those parts of evolution, where randomness produces a relatively large variability and, therefore, hermeneutical sciences could survive.

Advantages of the Stochastic Approach

The Bernoulli Space includes explicitly all sources of uncertainty and constitutes a macroscopic, holistic and abstract stochastic model for the uncertainty of real world. It makes it possible to transform any available quantitative/qualitative knowledge into a quantified model. The stochastic model can therefore be improved in principle by means of each additional observation and, thus, remains always up-to-date.

Prediction procedures constitute the basis of all stochastic procedures implying that stochastics can be build up as a unified theory for systematically investigating the elements of uncertainty.

The unified theory allows to draw up a complete list of desirable procedures for prediction and measurement purposes.

The quality of a stochastic procedure may readily be determined by means of the known reliability level and the given accuracy.

The Book of Change and Stochastic Science

In this context it is worthwhile to look at other approaches to develop a cosmology than that of modern science. The Chinese classic I Ging or *Yijing* (Book of Change)[54], which goes back to the early Zhou dynasty (about 1000 BC), is based on a world-view which seems to be close to that of Stochastic Science. It is based on the two concepts Yin and Yang which constitute the fundamental concepts of the Chinese world-view. They represent the primal opposing but complementary forces that are inherent in all things of the universe and which may smoothly change from one to the other.

Figure 1: The Yin Yang symbol representing “a contrastive relationship”.

The Book of Change describes a formal system which may be used for decision making. The system is based on two complementary types of lines[55], an unbroken solid line and a broken line with a gap in the center. These lines are used to build 64 hexagrams where each hexagram is a figure of six of these lines. The arrangement of the lines is determined by a random process, traditionally by a complex random drawing process of sound and broken yarrow stalks.

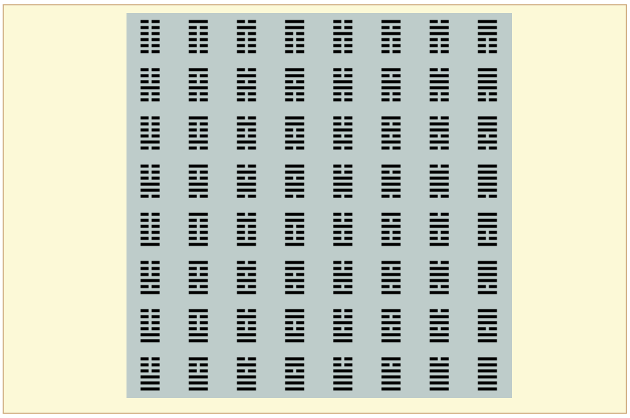


Figure 2: The 64 hexagrams of the Yijing.

Decisive for the interpretation is the arrangement of the lines within the hexagram. The interpretation and hence the statements depend on

- the hexagram as such,
- the position of each single line within the hexagram, and
- a number between six and nine, which is assigned to each line during the drawing process.

The number assigned to each line quantifies the degree of its stability, i.e., the strength of tendency to change to the complementary line.

Each hexagram stands for one human situation where the lines represent the different aspects of the situation and the numbers describe the tendency, or in other words, the degree of certainty that the aspect change to its complement.

[54]For details see: Karl Wulff (2006): *Naturwissenschaften im Kulturvergleich.* Harry Deutsch, Frankfurt, pp. 91.
[55]The line represent the two complementary principles *Yang* and *Yin*.

As stated by K. Wulff, the entire system of hexagrams represents a concept to describe the laws of evolution which govern not only mankind, but the entire universe. Mankind represents the microcosm within the macrocosm, where both are related to each other. Each hexagram and hence each human situation contains the possibility to change to one of the 63 other states implying that a human situation should not be looked at as isolated, but with regard to possible changes.

The Book of Change constitutes a rational system that is based on the assumption of a dynamic evolution and that aims at providing mankind some insight into the complex universal connectivity. The Book of Change reflects the knowledge that the universal connectivity produces fixed laws. There is an essential difference between the Chinese approach represented by the Book of Change and the Western approach as represented by modern science. In contrast to modern science, the Book of Change is not based on the belief in causal relations, but on what is called here stochastic relations. The same situation might lead to different developments according to the underlying tendencies which are quantified by the numbers six to nine assigned to each line of a hexagram.

Comparing the approach described in the Book of Change with Stochastic Science reveals fundamental similarities. Both approaches are based on

- universal connectivity,
- stochastic relations, and
- continual changes.

The difference refer to the technical details. The book of Change assumes the existence of a (universal) random experiment that reflects the universal connectivity and can be applied to reveal the relations in any given human situation. Moreover, the situation and especially the procedures' outcomes are quantified only partially. In contrast Stochastic Science develops for each given situation a specific and fully quantified process, which enables to answer the important questions about the procedures' reliability and accuracy. Such questions cannot adequately be dealt with in the Yijing system. Compared with modern science, however, the Yijing system appears to be rational and represents a formidable progress away from an irrational belief in ghosts and devils which probably was its starting point.

Examples

The following examples refer to the random variables considered in the examples of the previous learning units.

1. **Murder on the Orient-Express**

   Clearly, the murder on the Orient-Express, which as shown in the previous learning units, can be an object of stochastic science is accessible for methods used in contemporary science, particularly for DNA-investigations or other forensic examinations.

   However, it should be clear that the interpretation of the results obtained by a forensic analysis should take into account adequately randomness that affects the evidence producing process and the applied forensic experiment itself. if the random effects are not considered adequately then doubtful evidence is produced and may lead to judicial errors.

2. **The Bowl Filled with Chocolate Balls**

Similar as in the case of the murder on the Orient-Express, also the trivial process of grabbing chocolate balls out of a bowl can be analyzed in Stochastic Science in contrast to conventional science which is based on the Hamiltonian Formalism (William Rowan Hamilton reformulated Newtonian mechanics and developed a mathematical formalism that appealing in view of its simplicity and generality). Stochastic science makes available the appropriate rules and methods for investigating this everyday process.

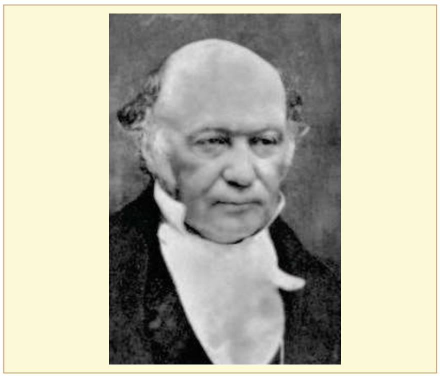


Figure 3: William Hamilton (1805 – 1865), Irish scientists.

Note, that the property that stochastic science may be applied for modeling any part of evolution is one of the most striking difference to contemporary science. Contemporary science yields useful results only in the cases that the random variability is sufficiently small and therefore can be neglected. Unfortunately, this is the case only in extremely rare cases and engineers have to add to the results obtained so-called safety margins as otherwise the products would not work or even be a danger.

Module 1.3: The Bernoulli Space

Content and Aim of the Module *The Bernoulli Space*

In the previous module the Bernoulli Space has been introduced as general stochastic model of the relation between past facts and future development. In this module the concept Bernoulli Space shall be further investigated. Different situations with respect to knowledge or ignorance are distinguished and some new terms and principles related to uncertainty are developed.

As a matter of fact the Bernoulli Space represents uncertainty about the future development of interest without assuming *ideal conditions* with respect to uncertainty as done in modern sciences, since Galileo Galilei formulated his famous Law of Falling Bodies. The Bernoulli Space as the stochastic model per se enables reliable and precise predictions and, thus, supports appropriate decisions.

Module 1.3 aims primarily at imparting the capability for deriving a Bernoulli Space in a given situation and for a given objective, where the Bernoulli Space is based on the complete, available knowledge, but not on postulates, belief or subjective opinions.

The second aim of this module is to show again that traditional science cannot really support understanding of the ongoing universal evolution. The artificial segmentation of the universal system prevents understanding the entire system. It leads necessarily to wrong decisions as can be seen of the present state of earth, which is characterized by a frightening physical and social disorder.

A successful passing of this module should enable to understand a given Bernoulli Space and to derive independently new Bernoulli Spaces for given situations. Moreover, understanding that everything and everybody has an impact on the future development should lead to a new sense of responsibility not only for the physical environment but also the entire human community.

Note that the Bernoulli Space constitutes *The Stochastic Model*, which must be necessarily used in order to describe the always existing ignorance of human beings and the inherent randomness of the evolution of universe.

Unit 1.3.1: The Deterministic Case

TARGET

Learning Unit 1.3.1 aims at illustrating the traditional scientific approach from view of the Bernoulli Space. The learner shall understand the limitations of deterministic models compared with stochastic models and realize that determinism prevents a deeper understanding of evolution.

CONTENT

The Bernoulli Space in a Deterministic World

In a deterministic world there is no randomness and no variability within evolution. The future development is completely determined by the facts, i. e., by the past. Thus, there is no principle difference between past and future. Future is simply a transformation of past.

There is no randomness in a deterministic world and no probability. Therefore, the image of the variability function $\mathcal{X}$ for any singleton $\{d\}$ is again a singleton $\{x(d)\}$ and the image of the random structure function degenerate to a set which contains only one element and this element is a so-called one-point probability distribution, where the event $\{x(d)\}$ for given d occurs with probability 1.

Uncertainty about the future development is solely generated by ignorance about the initial conditions given by the ignorance space $\mathcal{D}$.

Similar as in the case of a Bernoulli Space, the model of a process in the deterministic world refers to a pair of variables (X, D), where X represents the aspect of interest in the future and D the relevant aspects of the initial conditions. However, the future value of the variable X is not indeterminate, but is uniquely fixed as $x(d)$ by the actual value d of the deterministic variable D.

In a deterministic world the Bernoulli Space $\mathbb{B}_{X,D}$ has the following components:

$$\begin{aligned}
&\mathcal{D} = \{d\} \\
&\mathcal{X}(\{d\}) = \{x(d)\} \\
&\mathcal{P}(\{d\}) = P_{X|\{d\}}
\end{aligned}$$

$$\text{with } P_{X|\{d\}}(\{x\}) = \begin{cases} 1 & \text{for } x = x(d) \\ 0 & \text{for } x \neq x(d) \end{cases}$$

Generally, when formulating deterministic laws, knowledge about the initial conditions is assumed and, hence, the ignorance space reduces to a singleton $\mathcal{D} = \{d\}$.

Consequently, in a deterministic world it is sufficient to specify the function $x(d)$, which maps the past given by d onto the future given by $x(d)$. Once the function $x(d)$ is known, the entire evolution from its beginning to its end is in principle established.

Most of the models and Natural Laws in contemporary science assume a deterministic world and are of the above given form, despite the fact that none of the predictions made by means of such a model can actually be observed.

Modern Science excludes (with some tiny exceptions) a stochastic evolution and is founded on the postulate of a deterministic world neglecting all empirical experiences. Consequently, the Natural Laws do not allow a stochastic evolution. In contrast *Stochastic Science* does not exclude a deterministic evolution, since all our experiences are not enough evidence for allowing to exclude determinism with certainty. The ignorance about the nature of evolution is expressed in *Stochastic Science* by admitting both a deterministic and a stochastic evolution. As shown above a stochastic models includes both possibilities and, therefore, *Stochastic Science* gets by without the necessity to rely on belief.

However, the Bernoulli Space reveals that deterministic models are at the edge of possible models representing pathological or degenerate cases.

Deterministic World and Reality

A deterministic world is based on the assumption that the evolution is completely determined and fixed by (divine) 'Providence' for all eternity. Clearly, mankind will never be able to prove this assumption, but this it not at all decisive. All experiences of man with evolution contradicts this assumption and even physical science had to give up this assumption in certain areas of evolution.

All the more it is surprising that mankind does not abandon the conception of determinism neither in science nor in the individual realms. Whenever a problem occurs, i. e., an effect is observed, a search for 'the cause' is started, in order to eliminate the cause and, thus, prevent its future occurrence. By proceeding like this the universal connectivity which leads to a complex interplay of all existing things is neglected. Generally, mere symptoms are interpreted as 'causes' and, again very often, the reasons for declaring something as 'cause' have hardly anything to do with the observed effect.

The decomposition of evolution in cause-effect-chains which are out of touch with reality prevents a deeper understanding of evolution and opens the floatgates to quacksalvers and hidden persuaders. In all times including the present time cruelties and wars are started and founded on ridiculous causes. History shows that mankind follows willingly such evidently false argumentation, probably because humans are trained to think in causalities from their birth until they pass away.

Natural Laws, Determinism and the Bernoulli Space

Most of the Natural Laws formulated in physics are of deterministic type. In *Wikipedia, the free encyclopedia* the following features of physical laws are listed under the entry *Physical law* citing two sources[56]. Physical laws are:

- true (a.k.a.[57] valid). By definition, there have never been repeatable contradicting observations.
- universal. They appear to apply everywhere in the universe. (Davies)
- simple. They are typical expressed in terms of a single mathematical equation. (Davies)
- stable. Unchanged since first discovery, although they may have been shown to be approximations of more accurate laws.

[56] Paul Davies (1992): *The Mind of God*, ISBN 0.671-79718-2, and Richard Feynman (1965): *The Character of Physical Law*, ISBN 0-679-60127-9.

[57] also known as

- eternal. They appear unchanged since the beginning of the universe (according to observations). It is thus presumed that they will remain unchanged in the future. (Davies)
- omnipotent. Everything in the universe must comply with them (according to observations). (Davies)
- generally conservative of quantity. (Feynman)
- often expressions of existing homogeneities (symmetries) of space and time. (Feynman)
- typically theoretically reversible in time (if non-quantum), although time itself is irreversible. (Feynman)

If this list is meant seriously, it is revealing physical laws as fraud. A physical law given by a simple mathematical equation can never represent the universal connectivity and, therefore, will never be true or valid, universal, omnipotent or eternal. It appears that physicists are drunken by the 'intellectual beauty' of their mathematical derivations and cannot see reality anymore.

The attempt to derive from the physical laws the corresponding Bernoulli Spaces will readily show that most of the physical laws are either trivial or wrong statements.

EXAMPLES

1. **The Universal Law of Gravitation**

Figure 1: Sir Isaac Newton.

Sir Isaac Newton who has been regarded for almost 300 years as the founder of modern physical science, published in 1687 his famous Law of Gravitation, which states that every particle in the universe attracts every other particle with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. In other words, every particle is related to every other particle demonstrating the universal connectivity.

The mathematical formulation of the Law of Gravitation is given by:

$$F = g\,\frac{m_1 \cdot m_2}{r^2}$$

where g is a universal constant, m_1 and m_2 the masses of the two particles and r the distance between the two particles.

Newton's law has been refuted by now, but nevertheless it was looked upon as a Universal Law of Nature for a long time. Clearly, the law does not describe reality, because an isolated system with only two particles does not exist in this universe. By the way this statement follows from the law. The law describes a certain tendency of the two particles, but applying it in a special case would be tantamount to obtain with certainty a wrong result.

The law gives no room for the always existing ignorance in a special situation and does not allow to analyze the observed variability for making reliable and precise predictions on the one hand and to learn more about the special features of a given situation.

2. **The Ideal Gas Law**

The ideal gas law describes the state of an ideal gas. It says that the state of an amount of gas is determined by its pressure, its volume and its temperature. The mathematical model is as follows:

$$p\,V = n\,r\,T$$

where

$p =$ pressure
$V =$ volume
$n =$ number of moles of gas
$r =$ gas constant
$T =$ temperature

Similar as in the case of the Law of Gravitation the Law of Ideal Gas describes a tendency and, in fact, rather roughly. In any case the obtained values are never observed and whether they are useful or not depends on the situation and the aim of the investigation. A law, which yields results which are less far away from reality is the so-called Van der Waals Equation, but even this equation is only an approximation which cannot be used for making reliable predictions.

3. **Speed of Light**

One of the postulates which Albert Einstein needed for developing the special relativity states that the speed of light is in the empty space (which obviously does not exist) always the same and which is independent of the motion of the emitting body.

Einstein's postulate were necessary for developing mathematically further hypotheses, which changed the entire cosmology. However, one should be aware that Einstein derivations were not based on observations or experience, but rather on some puzzles.

In order to prove the statement one must exclude the contrary with certainty, which is, of course, not possible as humanity has not the possibility to control the velocity of photons. However, it is also impossible to disprove the statement because of the same reason. Thus, the existing ignorance is overcome be subjective belief, which as we have seen, makes any founded statement about the achieved reliability impossible.

4. **Schrödinger Equation**

The famous Schrödinger wave equation is a mathematical model describing the evolution of a wave. For a single-particle system let each eigenstate be denoted by $|r\rangle$, where r is a position vector and $|r\rangle$ is interpreted as the state of the particle at position r. The wavefunction is 'defined' as the projection of the state vector of a $|\psi(t)\rangle$ on the position basis and is given by:

$$\psi(r,t)\rangle \equiv \langle r|\psi(t)\rangle$$

By normalization one obtains

$$\int\limits_{\Omega} |\psi(r,t)|^2 d^3R = 1$$

where Ω denotes the entire space. Because of the normalization, the absolute square of the wavefunction is a number between 0 and 1. Hence it may be formally interpreted as a probability. However, the corresponding event is difficult to identify. According to the Copenhagen interpretation, the event refers to finding the particle in the infinitesimal region of volume d^3r around the position r.

The derivation of the wavefunction has been done completely mathematically based on postulates without reference to reality. Moreover, the wavefunction is completely deterministic and no randomness disturbs the evolution of the imaginary world of Schrödinger's wavefunction. The interpretation of the absolute square of the wavefunction as a probability has no real meaning and is primarily based on the fact that the integral over the whole space had been normalized.

The main problem with all these mathematical derivation is that they do not start with experience, but with a postulated truth in which one has to believe just as in the case of the divine truths in scholastic science. The rest is mathematics and obscure interpretations. This last claim shall be illustrated by the well known thought-experiment "Schrödinger's Cat".

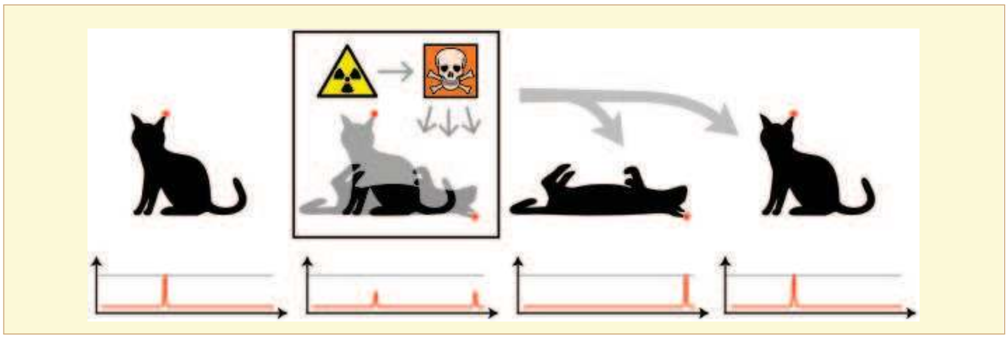


Figure 2: The thought experiment Schrödinger's Cat.

A cat, a flask containing a poison and a radioactive source are sealed in a box shielded against environmentally induced quantum decoherence. If an internal Geiger counter detects radiation, the flask is shattered, releasing the poison that kills the cat. The Copenhagen interpretation of quantum mechanics implies that after a while, the cat is simultaneously alive and dead. However, if we open the box, the cat is either alive or dead, not both alive and dead.

Unit 1.3.2: The Case Without Ignorance

TARGET

Scientific investigations of gambling led to the development of probability theory and the scientific handling of randomness. This learning unit aims at showing that gambling does not represent the true nature of randomness, but is much more related to ignorance.

CONTENT

The Bernoulli Space in Case of Complete Knowledge

Clearly, no ignorance or complete knowledge is a special, rare case and strictly speaking not possible, because of the universal connectivity. However, historically this case has played and plays a decisive role in science and caused in some sense many erroneous developments in dealing with uncertainty and its quantification. No ignorance means possess the truth and actually this is the dream and the aim of scientists.

No ignorance refers to the deterministic variable D and means that the ignorance space $\mathcal{D}$ contains only the true value, say d_0. In this case the domains and co-domains of the variability function $\mathcal{X}$ and the random structure function $\mathcal{P}$ are singletons, too. The co-domain of the variability function contains only one set namely the range of variability of $X|\{d\}$, and the co-domain of the random structure function only one probability distribution namely that of $X|\{d\}$. Thus, the Bernoulli Space for the pair of variable (X, D) can be represented by a singleton, a finite set and a set containing only one probability measure:

$$\mathbb{B}_{X,D} = \big(\{d_0\}, \mathcal{X}(\{d_0\}), P_{X|\{d_0\}}\big) \tag{17}$$

In this case randomness is the exclusive source of uncertainty. The stochastic nature of the situation modelled is evident by the set $\mathcal{X}(\{d_0\})$ and the corresponding probability distribution $P_{X|\{d_0\}}$.

If a deterministic cause-effect relation were to be assumed, $\mathcal{X}(\{d_0\})$ would shrink to a singleton, say $\{x_0\}$, and $P_{X|\{d_0\}}$ would degenerate. Thus, deterministic relations – as generally used in physics – are also covered by the approach adopted here, but emerge as more or less trivial and degenerated edge cases.

Figure 1 illustrates a notional example of a complete knowledge univariate case, where $\mathcal{X}(\{d_0\})$ is discrete.

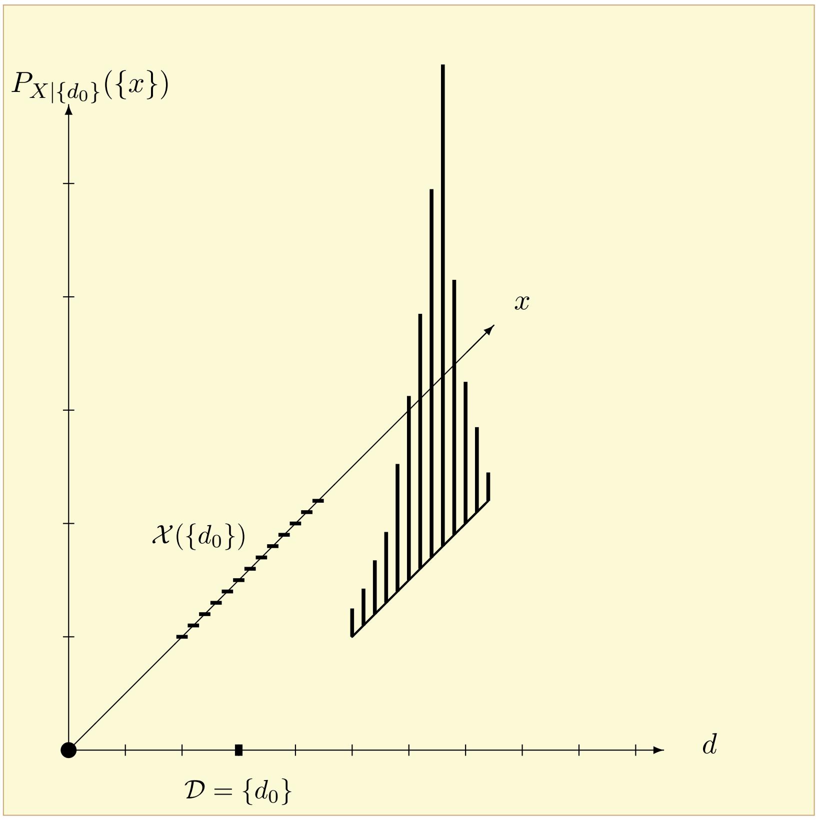


Figure 1: Graphical illustration of a Bernoulli Space in case of complete knowledge.

Gambling and Uncertainty

The process of quantification of uncertainty started with the quantification of gambling during the late Middle Ages leading to the famous correspondence between Blaise Pascal and Pierre de Fermat, which is regarded as the beginning of probability theory, although the word "probability" does not occur within the correspondence.

Assuming that the dice used for gambling are not loaded leads immediately to the fact that each possible outcome has the same probability. Or in other words the dice are purposefully produced symmetrically in such a way that none of the possible outcomes, when throwing them, is preferred. In gambling the situation with respect to the future development is completely known, as soon as the set of possible outcomes are known, because the number of outcomes determines completely the random structure, i.e., the probability distribution. For instance, when throwing two dices, the set of outcomes is given by:

$$\mathcal{X}(\{d_0\}) = \{(1,1),(1,2),\ldots,(2,1),(2,2,\ldots,(6,6)\}$$

implying that there are 36 equi-probable possible outcomes and, hence:

$$P_{X|\{d_0\}}(\{(i,j)\}) = \frac{1}{36} \quad \text{for } i = 1,2,\ldots,6 \text{ and } j = 1,2,\ldots,6$$

Consequently, gambling represents a situation in which the man-made evolution does not prefer any of the possible outcomes. This situation leads to maximum uncertainty and is therefore called chaos. In contrast to this man-made chaos, natural randomness always prefers certain outcomes and therefore may be looked upon as a principle of order. If each possible outcome occurs with equal probability, each outcome must be regarded likewise, just as in the case of ignorance. This similarity has led to the conclusion that randomness means disorder and represents blindness. Further more, randomness and ignorance were not clearly distinguished, but randomness was equated with ignorance. Clearly, this equalization supported the prevailing determinism and was immediately accepted.

Thus, the scientific occupation with randomness and uncertainty did not foster the abandonment of determinism, but even strengthen its position.

EXAMPLE

In the 1st and 3rd part of his masterpiece *Ars conjectandi* Jakob Bernoulli deals with gambling and, thus, with the case without ignorance. Only in the 4th part of his master piece *Ars conjectandi*, Jakob Bernoulli turns to the case with ignorance which necessitates to develop a measurement procedure. According to his own statements, he spent 20 years of research to this question and he judged the result as the major achievement of his life. The following examples are taken from the third part.

Figure 2: Jakob Bernoulli (1655 – 1705.

1. *Somebody puts a white and a black ball in an urn and promise a reward to three players A, B, C for the one who draws first the white ball. Each player has one draw. A should start, followed by B and finally followed by C. If none of the players draws the black ball, then none of the three gets the reward. After each draw the ball drawn is replaced in the urn. What are the odds of the three players?*

   The aspect of interest is the outcome of the draws of the players A, B, C, as this sequence determines who will get the payoff. Thus, the random variable is given by

   $$X = (X_A, X_B, X_C)$$

   where

   $$X_A = \begin{cases} 1 & \text{for} \quad A \text{ draws a black ball} \\ 0 & \text{for} \quad A \text{ draws a white ball} \end{cases}$$

   Let the random variable X_B and X_C be defined analogously. The deterministic variable D stands for the set of rules, including the number of balls in the urn, the number of players, the order of draws, the number of draws of each player and the rule of payoff. Because the rules imply the case of no ignorance, we omit a detailed definition of D and $\mathcal{D} = \{d\}$.

   The domain of the variability function is the singleton $\{d\}$, while the co-domain, which consists of a system of subsets, is also given by a singleton, namely by the set of possible outcomes of $X|\{d\}$:

   $$\begin{aligned} \mathcal{X}(\{d\}) \quad = \quad & \{(1,1,1),(1,1,0),(1,0,1),(0,1,1), \\ & (1,0,0),(0,1,0),(0,0,1),(0,0,0)\} \end{aligned}$$

   The range of variability $\mathcal{X}(\{d\})$ of $X|\{d\}$ determines the corresponding probability structure function:

   $$\mathcal{P}(\{d\}) = P_{X|\{d\}}$$

   with

   $$P_{X|\{d\}}(\{(i,j,k)\}) = \frac{1}{8} \quad \text{for } i,j,k = 0,1$$

The starting player A will win the payoff, if he draws the white ball. There are four corresponding outcomes, which represent the event E_A that player A wins:

$$E_A = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1)\}$$

Player B is the second one to draw a ball. He wins if he draws a white ball and A has not drawn a black ball. Thus, player's B winning event E_B is given by:

$$E_B = \{(1,0,0),(1,0,1)\}$$

The third player C wins, if he draws a white ball and player A and B have drawn each a black ball implying that the winning event E_C is a singleton:

$$E_C = \{(1,1,0)\}$$

Finally, the stakeholder wins if none of the players has drawn a white ball. His winning event E_S is therefore given by:

$$E_{St} = \{(1,1,1)\}$$

The probabilities of the winning events for the four participants in the game are as follows:

$$P_{X|\{d\}}(E_A) = \tfrac{4}{8} = \tfrac{1}{2}$$

$$P_{X|\{d\}}(E_B) = \tfrac{2}{8} = \tfrac{1}{4}$$

$$P_{X|\{d\}}(E_C) = \tfrac{1}{8}$$

$$P_{X|\{d\}}(E_{St}) = \tfrac{1}{8}$$

2. *Six person A, B, C, D, E and F take part in a game, where the latter persons are preferred by the stakeholder: First A and B play against one another, the winner plays against C. Again the winner plays against D and so on until the last player F is reached. The winner of the last play gets the payoff. In each play the two players have equal chance of winning. What are the odds of the three players?*

   Each player plays according to the above given order against each of the other players leading in all to $2^5 = 32$ different plays. Each game has two outcomes coded as 0 and 1, where 0 stands for the success of the player with alphabetically smaller order and 1 for the success of the player with alphabetically higher order. The random variable describes the outcome of the five plays, i. e. $X = (X_1, \ldots, X_5)$ with range of variability given by:

$$\mathcal{X}(\{d\}) = \{(i_1, i_2, \ldots, i_5) \,|\, i_j = 0, 1 \text{ and } j = 1, 2, \ldots, 5\}$$

   with $2^5 = 32$ equiprobable outcomes.

   Let E_A, E_B, E_C, E_D, E_E and E_F be the winning events for the five players. Then we have:

$$\begin{array}{rcl} E_A & = & \{(0,0,0,0,0)\} \\ E_B & = & \{(1,0,0,0,0)\} \\ E_C & = & \{(0,1,0,0,0),(1,1,0,0,0)\} \\ E_D & = & \{(0,0,1,0,0),(0,1,1,0,0),(1,0,1,0,0),(1,1,1,0,0)\} \\ E_E & = & \{(0,0,0,1,0),(0,0,1,1,0),(0,1,0,1,0),(1,0,0,1,0), \\ & & (0,1,1,1,0),(1,0,1,1,0),(1,1,0,1,0),(1,1,1,1,0)\} \\ E_F & = & \{(0,0,0,0,1),(0,0,0,1,1)(0,0,1,0,1),(0,1,0,0,1), \\ & & (1,0,0,0,1),(0,0,1,1,1),(0,1,0,1,1),(0,1,1,0,1), \\ & & (1,0,0,1,1),(1,0,1,0,1),(1,1,0,0,1),(0,1,1,1,1), \\ & & (1,0,1,1,1),(1,1,0,1,1),(1,1,1,0,1),(1,1,1,1,1)\} \end{array}$$

The probabilities for the different winning events are easily obtained by determining the size of the events:

$$P_{X|\{d\}}(E_A) = P_{X|\{d\}}(E_B) = \frac{1}{32}$$

Thus, player A and B have the same winning probability.

Player's C winning probability is given by:

$$P_{X|\{d\}}(E_C) = \frac{2}{32} = \frac{1}{16}$$

For player D two of his five plays are irrelevant and, therefore his winning event consists of four elements:

$$P_{X|\{d\}}(E_D) = \frac{4}{32} = \frac{1}{8}$$

For player E three of his five plays are irrelevant and, therefore, his winning event consists of eight elements:

$$P_{X|\{d\}}(E_E) = \frac{8}{32} = \frac{1}{4}$$

Finally, for player F only one play is relevant and, therefore, his winning event consists of sixteen elements:

$$P_{X|\{d\}}(E_F) = \frac{16}{32} = \frac{1}{2}$$

Only the players A and B have equal odds. The odds for the other players are doubled from C to F.

Unit 1.3.3: The General Case

TARGET

Learning Unit 1.3.3 shall illustrate the necessity of specifying ignorance within a stochastic model, because ignorance is a characteristic feature of mankind and the main source of uncertainty about the future development. Specifying ignorance opens the possibility to learn systematically in order to reduce it. Otherwise, the learning process meanders and gets generally inefficient.

CONTENT

Introduction

The general case is characterized by ignorance, which is also called "incomplete knowledge". It means that the values of some relevant facts of the initial conditions represented by a deterministic variable D are unknown - this is the situation most commonly encountered in practice. It is clear from what has gone before, that in this case one should not assume that the value of d of D is known, but admit all those values of D which cannot be excluded in the given situation and which form the ignorance space $\mathcal{D}$.

Consequently, the set $\mathcal{D}$ is no longer a singleton implying that the image sets of the two functions $\mathcal{X}$ and $\mathcal{P}$ contain possibly also more than one element. The larger the set $\mathcal{D}$, the larger the degree of ignorance. Reversing this logic, it now seems intuitively reasonable to suggest that the existence of partial knowledge will ensure that $\mathcal{D}$ is at least bounded.

The concept of 'ignorance' has been introduced as lack of knowledge concerning the true value of the initial conditions represented by the deterministic variable D. The given process determines the variability function as well as the random structure function.

Consider for example the situation of repeating independently a Bernoulli experiment with the number of successes being the aspect of interest. Then the type of the corresponding variability function as well as the random structure function are determined by the given process arrangement.

The above example refers to the case that the type of the variability function and the type of the random structure function are completely known. In general, this is not the case and the variability function as well as the random structure function are not known and have to be specified by exclusion procedures similar to that for selecting the ignorance space.

Selection of the Variability Function $\mathcal{X}$

The images $\mathcal{X}(\{d\})$ for $d \in \mathcal{D}$ are selected in the same way as the ignorance space itself. Any value x which is out of the question is exclude. Those values x which cannot be excluded in the given situation form the range of variability $\mathcal{X}(\{d\})$ to be selected.

It is of great importance not to exclude values which in the given situation may occur. Therefore, only those values may be excluded which in fact cannot occur.

Any available knowledge (about *what is not*, should be used to select the range of variability $\mathcal{X}(\{d\})$ for $d \in \mathcal{D}$. However, it is critical that only confirmed knowledge be exploited in this manner. Using unconfirmed knowledge may lead to a set $\mathcal{X}(\{d\})$ which no longer contains all the possible outcomes of $X|\{d\}$.

Selection of the Random Structure Function

The images of the random structure function $\mathcal{P}$ are probability measures, where each probability measure is a member of a certain family of probability distributions which are introduced in Learning Unit 1.2.7. The existing partial knowledge or experience about the given process generally allows to exclude most of the families of probability measures by knowing, for instance, that the frequency of values x is decreasing (or increasing) with increasing x. For other processes it is known that the frequency of values x is decreasing with the distance $|x - x_0|$, where x_0 might be known or unknown.

Selection of the family of probability distributions should be based on the knowledge about the initial conditions and on available knowledge about the process in question. However, it should not be based on a given set of observed data and a best fit criterion, because the data are the outcome of randomness and fitting the probability distribution to randomly generated data may lead to a nonconforming result.

Complete Ignorance

Situations involving complete ignorance about *what is not* are not amenable to quantitative analysis - one must wait until sufficient knowledge about the process on hand is available in order to have the chance to describe it in an useful way. This leads to the question of the required minimum amount of knowledge, which will be answered later. The fact that situations with complete ignorance are not treated, does not diminish the real–world applicability of the scientific approach being developed here. In real world, when there is no prior knowledge, only 'inspiration' can be employed - and inspiration is not amenable to quantitative modeling!

Graphical Representation

A graphical illustration of the Bernoulli Space for a notional incomplete knowledge situation is provided in Figure 1 below. Here, the deterministic variable D takes one of four possible values, i. e., $\mathcal{D} = \{d_1, d_2, d_3, d_4\}$. For each value d_i of D, there is the range of variability $\mathcal{X}(\{d_i\})$ and the corresponding probability distribution $P_{X|\{d_i\}}$ of the random variable $X|\{d_i\}$. As was the case with the example in the preceding Learning Unit 1.3.2, the random variable X is again of discrete type.

The overall range of variability of X is given by

$$\mathcal{X}(\mathcal{D}) = \bigcup_{i=1}^{4} \mathcal{X}(\{d_i\})$$

Two of the individual ranges; $\mathcal{X}(\{d_1\})$ and $\mathcal{X}(\{d_4\})$, are highlighted in Figure 1. They are disjoint. Therefore, if a value x of $X|\mathcal{D}$ is observed, it can be used to exclude one of the two values d_1, d_4 from $\mathcal{D}$ and thus reduce ignorance. This demonstrates, in a rather direct manner, how observations may be employed to reduce $\mathcal{D}$, and the overall range of variability for X.

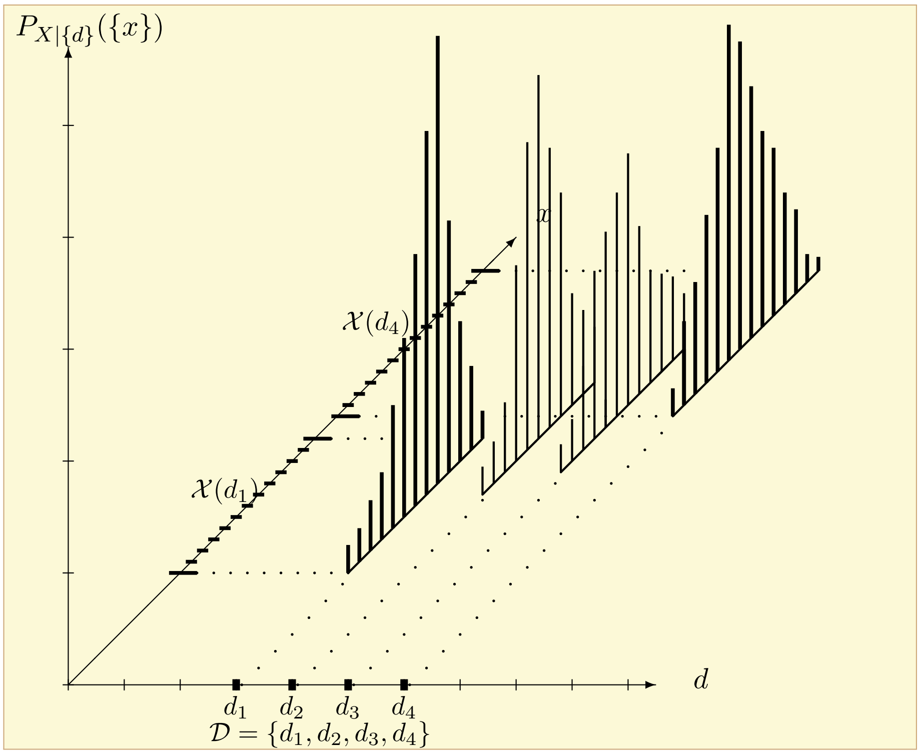


Figure 1: Graphical illustration of a Bernoulli Space in case of ignorance.

Necessity of Specifying the Ignorance

Besides the artificial case of gambling and some other constructed situations, where the initial conditions are designed in a way that none of the possible outcomes is preferred to others, there is always more or less large ignorance about the initial conditions and in most of these cases ignorance constitutes the major source of uncertainty. Therefore, it is of utmost importance to specify the existing ignorance, as otherwise it would not be possible to reduce it systematically.

Including ignorance explicitly in the stochastic model reflects strikingly the stochastic approach which is based on the knowledge that human beings will never be able to grasp *what is*, i. e., truth, but will be restricted to realize *what is not*. Thus, there will be always a more or less large amount of ignorance.

Examples

1. **Drawing without Repetition from a Set**

   Consider a set of $N = 10$ elements with d elements being of Type A and $10 - d$ elements being of Type B. The process consists of randomly[58] selecting by one draw $n = 4$ elements out of the set, and the aspect of interest is the future number of elements of Type A among them.

   Thus, we obtain the following variables:

   - Random variable:
     X = number of Type A elements among those selected.
   - Deterministic variable D:
     The probability distribution of X is completely determined by the known value of

[58]The meaning of "randomly" is that each subset of four elements has the same probability to be selected.

the overall number of elements, $N = 10$, the known number of elements to be drawn, $n = 4$ and the unknown value d of the elements of type A. Because the value of the overall number of elements and the number of elements to be drawn are known, it is not necessary to include them in the ignorance space. Therefore, the following deterministic variable D is obtained:

D = number of type A elements within the set.

If there is no knowledge at all about the value of D, we obtain the following ignorance space $\mathcal{D}$:

$\mathcal{D} = \{0, 1, \ldots, 10\}$

The variability function $\mathcal{X}$ is given by:

$\mathcal{X}(\{d\}) = \{\max_{d\in\mathcal{D}}(0, 4-(10-d)), \max_{d\in\mathcal{D}}(0, 4-(10-d))+1, \ldots, \min_{d\in\mathcal{D}}(d, 4)\}$

The random structure function $\mathcal{P}$ is given for $d \in \mathcal{D}$ by:

$\mathcal{P}(\{d\}) = P_{X|\{d\}}$ with $P_{X|\{d\}} \sim H(10, d, 4)$, where H denotes the Hypergeometric Probability Distribution, i. e., for $x \in \mathcal{X}(\{d\})$:

$$P_{X|\{d\}}(\{x\}) = \frac{\binom{d}{x}\binom{10-d}{n-x}}{\binom{10}{4}}$$

2. **Drawing by Repetition from a Set**

Consider the same situation as described in the previous example, except for the mode of drawing. In the previous example all the $n = 4$ elements were selected by one draw. Now, only one element is drawn, but the process is repeated four times, where repetition means that the original situation is restored before each of the draws. In this case we get the same pair of variable as in the previous example:

- Random variable:
  X = number of Type A elements among those selected.
- deterministic variable:
  D = fraction of type A elements within the set.

If there is no knowledge at all about the value of D, we obtain the following ignorance space $\mathcal{D}$:

$\mathcal{D} = \left\{0, \frac{1}{10}, \ldots, \frac{9}{10}, 1\right\}$

The variability function $\mathcal{X}$ is given by:

$$\mathcal{X}(\{d\}) = \begin{cases} \{0\} & \text{for} \quad d = 0 \\ \{0, 1, \ldots, 10\} & \text{for} \quad 0 < d < 1 \\ \{10\} & \text{for} \quad d = 1 \end{cases}$$

The random structure function $\mathcal{P}$ is given for $d \in \mathcal{D}$ by:

$\mathcal{P}(\{d\}) = P_{X|\{d\}}$ with $P_{X|\{d\}} \sim Bi\left(4, \frac{d}{10}\right)$, where Bi denotes the Binomial Distribution, i. e., for $x \in \mathcal{X}(\{d\})$ with $P_{X|\{d\}}(\{x\}) = \binom{4}{x} d^x (1-d)^{4-x}$.

Unit 1.3.4: Uncertainty Space

TARGET

As demonstrated in the earlier learning modules, the source of human problems is the uncertainty about future developments. For solving the problems, reliable and precise predictions are necessary leading to the development of stochastic science. The identification of the amount of existing uncertainty is of great importance. This learning unit introduces the corresponding concept.

CONTENT

Two Important Classes of Tasks with Respect to Uncertainty

The stochastic approach is based on the pair of variables (X, D). As already indicated in Learning Unit 1.2.6, there are several general classes of problems that may be solved using the stochastic approach. Two of these classes, which correspond to the division of uncertainty have already been identified, namely the reduction of uncertainty related to X and reducing ignorance related to D, respectively.

The existence of the first problem class is already well established here. It relates to predicting the future value of a given process quantity of interest X. In order to make reliable and accurate *predictions*, the appropriate deterministic variable D, its values $d \in \mathcal{D}$ and the probability distributions $P_{X|\{d\}}$ must be considered. Any meaningful prediction for X is necessarily a subset of $\mathcal{X}(\mathcal{D})$.

The second problem class relates to discover or determinate an unknown, fact. To determine the unknown fact a process or 'experiment' must be designed and an appropriate random variable X must be selected that relates to the unknown fact represented by the deterministic variable D. Following general practice in science, procedures to determine a fixed, but unknown value are called measurements procedures.

Applying a measurement procedure makes only sense, in case the true value is not known and, hence, $\mathcal{D}$ contains more than one element. The general aim of a measurement procedure is to arrive at a smaller $\mathcal{D}$ and, thus, to reduce the degree of ignorance. Any meaningful measurement is necessarily a subset of $\mathcal{D}$.

The Uncertainty Space

The extent of the uncertainty (nurtured by ignorance and randomness) characterizing any real-world situation is quantified by

$$\mathcal{U}_{X,D} = \left\{(x, d) \,\middle|\, x \in \mathcal{X}(\{d\}),\, d \in \mathcal{D}\right\} \tag{18}$$

Irrespective of the problem class, the overall purpose is to reduce the size of the set (18) defined by the corresponding Bernoulli Space which will be referred to as the uncertainty-space.

Uncertainty has tow parents or two dimensions, namely ignorance and randomness. Ignorance features no structure and, therefore, represents maximum disorder or chaos. Randomness on the other hand stands in general for structure and, thus, for order. One of the most difficult problems in this context is the identification of the deterministic variable and the establishment of its relation to the random structure. Therefore, a representation of the deterministic variable D for a given random variable X with innately specified connection between D and X would be of great advantage and is derived in the subsequent learning unit.

Example

1. **Uncertainty Space for Drawing without Repetition**

   Consider the first example of the previous learning unit, i.e., from a set of $N = 10$ elements with d elements being of Type A and $10 - d$ elements being of Type B a set of $n = 4$ elements is randomly drawn with the aspect of interest being the number of elements of Type A among them.

   The uncertainty space for a given situation contains everything which cannot be excluded as initial condition as well as everything which might happen in the future. The uncertainty space for the above described process of drawing is given in Figure 1:

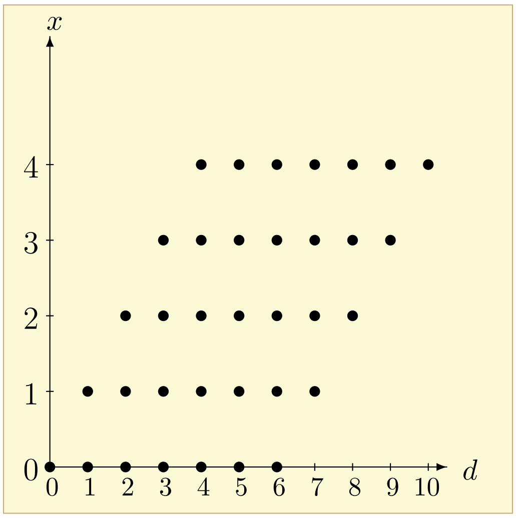


Figure 1: The uncertainty space for the process of drawing 4 elements out of the set of 10 elements with d elements being of Type A and $10 - d$ elements of Type B.

   Note that there is nothing new or surprising with regard to the quantitative description of the drawing process except for the fact that the existing ignorance is explicitly incorporated in and stated by the model. By this it becomes possible to take the existing ignorance into account when making a reliable prediction.

   For example, if it would be known that the number d of elements of Type A cannot be smaller than 5 and larger than 8, the uncertainty space would be as displayed in Figure 2.

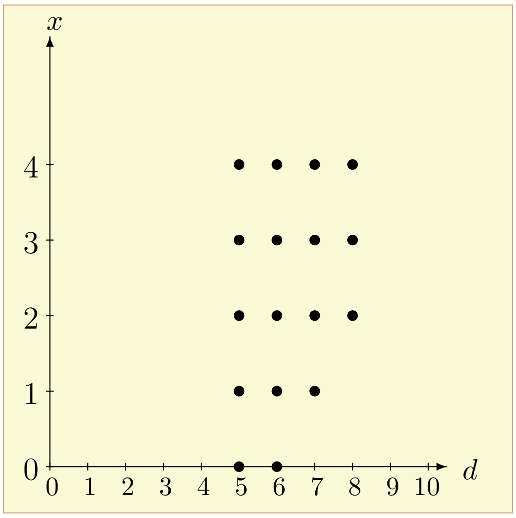


Figure 2: The uncertainty space for the process of drawing 4 elements out of the set of 10 elements when the number of elements being of Type A is not smaller than 5 and not larger than 8.

Unit 1.3.5: Representation of Probability Distributions

TARGET

One central topic in probability theory is the development of different families of probability measures. From a mathematical point of view a family of probability distributions should be as general as possible containing a multitude of different types of probability distribution. From a practical point of view such a general family is rather useless, as it cannot be assigned to a special type of situation.

A system of families of probability distributions is useful only if the members of the families can be distinguished by means of qualitative properties. This learning unit introduces a unified representation of probability distributions, which subsequently is used in Learning Unit 1.3.6 for defining the announced families.

CONTENT

There are numerous types of probability distribution proposed and investigated in probability theory. Unfortunately, this diversity is a major problem when selecting an appropriate probability distribution for a given situation. This is due to the fact that there is not a one to one correspondence between possible situations and available distributions as neither situations nor probability distributions are classified. Only in some rare cases, a probability distribution investigated in probability theory is uniquely determined by the process and the random variable X in question. Generally, for each situation there are many seemingly appropriate distributions and selection must be made using inadequate criteria.

Let the range of variability of the univariate random variable X be given by

$$\mathcal{X} = \{x_1, x_2, \ldots, x_N\} \quad \text{with } x_1 < x_2 < \ldots < x_N$$

Note that the range of variability of any real-world random variable has necessarily the above form.

The problem is to derive a generic representation of the corresponding probability mass function given by

$$f_X(x_i) = P_X(\{x_i\}) = p_i > 0 \quad \text{for } i = 1, \ldots, N \tag{19}$$

Having in mind that in many cases a continuous approximation of the realistically discrete probability distribution would be of a great use, a function $f : \mathbb{R} \to \mathbb{R}^+$ is needed, which is positive for any real x and which meets the relations given by (19).

It is easily seen that the following representation meets the requirements:

$$f_X(x) = e^{\sum\limits_{i=0}^{N} \lambda_i \, x_i} \tag{20}$$

where the coefficients λ_i are defined as solutions of the following equations:

$$\begin{aligned} f_X(x_i) &= p_i \qquad \text{for } i = 1, 2, \ldots N \\ \sum_{i=1}^{N} f_X(x_i) &= 1 \end{aligned} \tag{21}$$

With (20) a simple and universally applicable representation of a probability distribution in the univariate case is available and the problem can be addressed whether it is possible to decompose the total set of probability distributions into disjoint families each representing exactly one type of situation.

The Set of Probability Distributions

With (20) the following set of probability distributions over a range of variability $\mathcal{X}$ represented by the corresponding probability mass functions is obtained:

$$f_{\mathbb{P}_{\mathcal{X}}} = \left\{ e^{\sum\limits_{i=0}^{\infty} \lambda_i x^i} \;\Big|\; \sum_{x \in \mathcal{X}} e^{\sum\limits_{i=0}^{\infty} \lambda_i x^i} = 1, \lambda_i \in \mathbb{R} \right\} \tag{22}$$

Having derived a generic representation of a probability distribution, the problem has to be solved to find a partition of $\mathbb{P}_{\mathcal{X}}$ into families, where each family represents a real world situation with some characteristic properties.

EXAMPLES

1. **Throwing a Die**

   The aspect of interest when throwing a die is the thrown number. Hence:

   $$X = \text{ thrown number}$$

   The probability mass function is constant with the following representation:

   $$f_X(x) = \frac{1}{6} \qquad \text{for } x \in \{1, 2, \ldots, 6\} \tag{23}$$

   A different but equivalent representation is obtained as follows:

   $$f_X(x) = e^{\lambda_0} \qquad \text{for } x \in \{1, 2, \ldots, 6\} \tag{24}$$

   with $\lambda_0 = -\log 6$.

2. **General Case**

   Assume a random variable X together with a deterministic variable D with actual value d and with the probability distribution given by the probability mass function or in case of a continuous approximation by the density function:

   $$P_{X|\{d\}}(\{x\}) = f_{X|\{d\}}(x) \qquad \text{for } x \in \mathcal{X} \tag{25}$$

   or equivalently:

   $$P_{X|\{d\}}(\{x\}) = e^{\log\left(f_{X|\{d\}}(x)\right)} \qquad \text{for } x \in \mathcal{X} \tag{26}$$

   Expand $\log\left(f_{X|\{d\}}(x)\right)$as a Taylor series around $x = 0$, then for $x \in \mathcal{X}$:

   $$P_{X|\{d\}}(\{x\}) = e^{\log\left(f_{X|\{d\}}(0)\right) + \sum\limits_{i=1}^{\infty} \frac{\frac{d^i}{dx^i}\log\left(f_{X|\{d\}}(x)\right)\Big|_{x=0}}{i!} x^i} = e^{\sum\limits_{i=0}^{\infty} \lambda_i(d) x^i} \tag{27}$$

   with $\lambda_0(d) = \log\left(f_{X|\{d\}}(0)\right)$ and $\lambda_i(d) = \frac{\frac{d^i}{dx^i}\log\left(f_{X|\{d\}}(x)\right)\Big|_{x=0}}{i!}$ for $i = 1, \ldots$.

Unit 1.3.6: Families of Probability Distributions

Target

In probability theory there is a huge variety of probability distributions and for one and the same case a multitude of different probability distributions seem to be adequate. This learning unit shall order the possible distribution according to given situations aiming at assigning exactly one family for one type of situation.

Content

The main subject of investigation in Bernoulli Stochastics is the structure of variability in the outcome of real-world processes. Any classification of the structure for use in stochastics must, therefore, be based on properties describing possible patterns of variability. These patterns are expressed essentially by the curve of the probability mass function. From mathematical curve sketching it is known that a curve may be adequately described by its zero and its extremes. As a probability mass function has no zeros, number and types of extremes must be taken for describing the curve and for deriving an appropriate partition of the set of probability distributions into disjoint families.

Note that the information about the number of extremes is basically of qualitative nature and does not include quantitative information as, e.g., location and extent of the extremes. Note further that classifying the probability mass functions according to their extremes coincides with the purely mathematical classification according to the degree of the polynomials in the exponent of the density function in its generic representation.

Family of Constant Probability Distributions

The simplest case is characterized by the fact that no points of inclination or disinclination of X can be observed and, therefore, implying that the density function is constant.

In this case the range of variability $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ constitutes the only quantitative information necessary for determining "exactly" the probability distribution:

$$f_X(x_i) = \begin{cases} \frac{1}{N} & \text{for } i = 1, 2, \ldots, N \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

In view of the generic representation of f derived in the previous learning unit the following values of the distribution parameters are obtained:

$$\begin{aligned} \lambda_0 &= -\log N \\ \lambda_k &= 0 \quad \text{for } k = 1, 2, \ldots \end{aligned} \tag{29}$$

The conditions (29) on the values of λ_i define the *family of constant probability distributions* denoted by $\mathbb{P}_0$. Note that the probabilities are completely determined by the size of the range of variability $\mathcal{X}$. This fact indicates the decisive role of the range of variability.

Family of Monotonic Probability Distributions

There are real processes exhibiting the qualitative property that the smaller (or the larger) the value is the more often it is observed. In this case there is exactly one maximum either at the left or right bound of the range of variability and a minimum at the opposite side.

The above described qualitative property leads to random variables with monotone density functions with two cases "the smaller the more probable", and, "the larger the more probable". In the former, the exponent of the exponential probability mass function decreases monotonically, whilst in the latter it increases monotonically.

Thus, the *family of monotonic distributions* is divided into two sub-families given by:

$$\begin{aligned} f_X(x_1) &> f_X(x_N) \\ f_X(x_i) &\geq f_X(x_{i+1}) \quad \text{for } i = 1, \dots N-1 \end{aligned}$$

and, respectively,

$$\begin{aligned} f_X(x_1) &< f_X(x_N) \\ f_X(x_i) &\leq f_X(x_i) \quad \text{for } i = 1, \dots N-1 \end{aligned}$$

The entire family of monotonic probability distributions is denoted $\mathbb{P}_1$

Family of Uni-Extremal Probability Distributions

The next family is defined by the existence of exactly one extreme in the interior of the range of variability $\mathcal{X}$. Similar as in the case of monotonic distributions, there are two different sub-families, the one refers to an inner maximum, i. e., a peak, and the other to an inner minimum, i. e., a sink. Again the family defining property is a merely qualitative one, which might be deduced by knowledge about the process in question and, of course, by past observational experiences.

The two sub-families of uni-extremal probability distributions are formally defined by the following two sets of conditions:

There is exactly one $x_m \in \mathcal{X}$ with $1 < m < N$ and

$$\begin{aligned} & f_X(x_1) < f_X(x_m) > f_X(x_N) \\ & f_X(x_i) \leq f_X(x_{i+1}) \quad \text{for } i = 1, 2, \dots, m-1 \\ & f_X(x_i) \geq f_X(x_{i+1}) \quad \text{for } i = m, m+1, \dots, N-1 \end{aligned} \tag{30}$$

or, respectively,

$$\begin{aligned} & f_X(x_1) > f_X(x_m) < f_X(x_N) \\ & f_X(x_i) \geq f_X(x_{i+1}) \quad \text{for } i = 1, 2, \dots, m-1 \\ & f_X(x_i) \leq f_X(x_{i+1}) \quad \text{for } i = m, m+1, \dots, N-1 \end{aligned} \tag{31}$$

The entire set of uni-extremal probability distributions is denoted $\mathbb{P}_2$.

The variability of a very large number of random variables follows a uni-extremal probability distribution with one peak. This extremely important case is called uni-modal and denoted $\mathbb{P}_{2,1}$ with the mode being the value for which the peak is adopted.

Moreover, under rather general conditions the probability distribution of a sum of sufficiently many copies of an arbitrary random variable is uni-modal. Very often this property is obtained already for a rather small number of copies. Thus, one can state that the uni-modal probability distributions have paramount importance.

Family of Multi-Extremal Probability Distributions

The generalization to families with k relative extremes is straightforward. The family of k-extremal probability distributions for a random variable X with range of variability $\mathcal{X}$ is characterized by the following conditions:

There are exactly k different elements $x_{m_j} \in \mathcal{X}$, $j = 1, \ldots, k$ with $x_1 < x_{m_1} < \ldots < x_{m_k} < x_N$. For each k there are again two sub-families, with different requirements for k even or k uneven.

- k even:

$$\begin{array}{l} x_1 < x_{m_1} > x_{m_2} < \ldots > x_{m_k} < x_N \\ \text{monotonic progression between subsequent extrema} \end{array} \tag{32}$$

  or, respectively

$$\begin{array}{l} x_1 > x_{m_1} < x_{m_2} > \ldots < x_{m_k} > x_N \\ \text{monotonic progression between subsequent extrema} \end{array} \tag{33}$$

- k uneven:

$$\begin{array}{l} x_1 < x_{m_1} > x_{m_2} < \ldots < x_{m_k} > x_N \\ \text{monotonic progression between subsequent extrema} \end{array} \tag{34}$$

  or, respectively

$$\begin{array}{l} x_1 > x_{m_1} < x_{m_2} > \ldots > x_{m_k} < x_N \\ \text{monotonic progression between subsequent extrema} \end{array} \tag{35}$$

The entire set of k-extremal probability distributions is denoted $\mathbb{P}_k$ for $k = 3, 4, \ldots$.

Simple and Compound Random Variables

The above given families of probability distributions give reason to distinguish between simple and compound random variables or random processes.

If the probability distribution of a random variable belongs to the constant, monotonic or uni-modal family the random variable is called *simple* and otherwise *compound*.

A process of interest may be a simple or a compound process. A simple process is a one step process, which proceeds directly to the future terminal state. A compound process proceeds in at least two steps, where the properties of the second step may depend on the result of the first one.

The probability mass function of a compound random variable may have more than one inner extremum. In many cases the considered compound process proceeds in as many steps as there are maxima (or minima). The random structure of any step depends on the state reached by the preceding one, where the number of different states may determine the number of inner maxima or minima of the probability mass function.

EXAMPLES

1. **Income Distribution**

   Consider a given population, where each individual has an income. Let the process consists of selecting one person randomly, with the outcome of interest X being that person's income. For real-world populations, the type of distribution of X is well-known.

Figure 1: Vilfredo Pareto

The Italian engineer, sociologist, economist, and philosopher, Vilfredo Pareto discovered that the income in human populations is monotonically decreasing. Accordingly he proposed a power law probability distribution for modeling income which is known as Pareto distribution. The Pareto distribution is a continuous approximation given by the density function:

$$f_X(x) = \frac{\alpha x_{min}^{\alpha}}{x^{\alpha+1}} \text{ for } x > x_{min}$$

where x_{min} is the smallest possible income.

2. **Physical Measurements**

   Most physical measurement procedures are assumed to be processes generating a random variable X with a uni-modal probability distribution. Thus, the probability mass function is uni-modal, i.e., $f_X \in \mathbb{P}_{2,1}$.

3. **Drawing Randomly from Different Sets**

   The probability mass function has two or more maxima, if the random variable is based on different sub-processes which become active with certain probabilities. Consider for example the process of drawing one element from a set consisting of several subsets with different numerical representations. In this case there will be in general as many maxima of f_X as there are different subsets.

Unit 1.3.7: Generic Deterministic Variable

Target

Learning Unit 1.3.7 aims at developing a generic representation of the deterministic variable D, which is simple and nevertheless characterizes the random structure and, thus, the actual probability distribution.

Content

As already mentioned, for any realistic situation the range of variability $\mathcal{X}$ of a random variable X is finite and, hence, in particular bounded. This fact follows because of two reasons:

- Human resolution even of the best measurement devices is finite.
- The universe is must be considered as finite.

Therefore, in the here considered univariate case, the probability distribution P_X of X is completely determined[59] by the sequence

$$(\mu_1, \mu_2, \ldots, \mu_K) \tag{36}$$

of values of the moments $E\left[X^k\right]$, $k = 1, 2, \ldots, K$. The moments of a random variable have an easy to grasp meaning and it is generally not difficult to derive procedures for determining their actual values. Hence, the following generic representation of the deterministic variable D, for a given random variable X, is selected:

$$D = \Big(E[X], E[X^2], \ldots, E[X^K]\Big) \tag{37}$$

with actual value $d = (\mu_1, \mu_2, \ldots, \mu_K) \in \mathcal{D}$.

Assume that for $d = (\mu_1, \mu_2, \ldots, \mu_K) \in \mathcal{D}$ the corresponding probability distribution or equivalently the probability mass function

$$f_{X|\{d\}}(x) = e^{\sum\limits_{j=0}^{K} \lambda_j(d)\, x^j}$$

shall be determined. This is a purely mathematical problem consisting of finding the solution $(\lambda_0(d), \lambda_1(d), \ldots, \lambda_K(d))$ of the following system of equations, where $\mu_0 = 1$.

$$\sum_{x \in \mathcal{X}(\{d\})} x^k\, e^{\sum\limits_{j=0}^{K} \lambda_j(d)\, x^j} = \mu_k \quad \text{for } k = 0, \ldots, K \tag{38}$$

If a continuous approximation is used, the sums in the above equations are replaced by integrals.

[59]For a proof see Hausdorff, F., (1921). Summationsmethoden und Momentenfolgen. *Math. Z.* 9, 74-108, 280-299.

Example

1. **Production Process**

   Consider a manufacturing process of a technical product. Of interest is the question whether or not the product to be produced will be conforming with the specifications or not. If it is conforming to the specifications it can be delivered to the customer, otherwise it should not be delivered. In this case the random variable describing the aspect of interest in the future is a so-called indicator variable, which adopts the value 1 if the event of interest occurs and the value 0, if it does not occur.

   $$X = \text{ indicator variable for the event of producing a conforming item}$$

   Production processes are often very complicated and describing the actual relevant state is difficult if not impossible. In order to have the probability distribution completely defined in this simple case, the value of the first moment of X is needed. Hence, we obtain the following deterministic variable D:

   $$D = E[X] \tag{39}$$

   The value of the first moment $E[X]$ is just the success probability, i. e., the probability that the event of interest will occur.

2. **Technical Measurement Procedure**

   All technical measurement procedures operate according to the same principle. If the actual value m of a variable M shall be determined, a measurement experiment must be performed, where the probability distribution of a random variable X defined by the future outcome of the experiment depends more or less heavily of the value m to be determined.

   The probability distribution of almost all random variables X defined by a measurement experiment is uni-modal. The deterministic variable D to be selected is given by

   $$D = (E[X], E[X^2], \ldots, E[X^K]) \tag{40}$$

   where the values of the moments depend on the value m to be determined. Obviously, the problem of selecting an appropriate number K is of utmost importance and will be solved in the next Learning Unit 1.3.8.

Unit 1.3.8: Selection of Probability Distributions

Target

In statistics probability distributions are generally obtained by first selecting a family of distributions (often the family of normal distributions) and subsequently fitting the distribution to some observations. In this learning unit a new selection principle is introduced, which is closely related to the classification of probability distributions developed in Learning Unit 1.3.6.

Content

Introduction

In order to select an appropriate random structure function for a given random variable X in a meaningful way, the corresponding family of probability distributions should necessarily be known.

This knowledge is available, if number and type of extremes of the corresponding probability mass function have been identified by sufficient empirical experience or by theoretical insight in the process of interest enabling the exclusion of any other family.

If an identification of the family is impossible, then the description of the random structure must cover more than one family and, therefore, it is to be expected that the achieved results are more or less useless because of inaccuracy. In the following it is assumed that the available knowledge is sufficient to exclude all families of probability distributions except for the actual one.

Principle of Minimum Information

The question has to be answered how much quantitative information about the actual value of d of the deterministic variable D is necessary for selecting an appropriate probability distribution $P_{X|\{d\}}$ in a given situation with identified family of probability distributions $\mathbb{P}_k$.

Clearly, the range of variability $\mathcal{X}(\{d\})$ of $X|\{d\}$ must be given in any case, because, if the possible outcomes of $X|\{d\}$ are not given, it is impossible to assign probabilities to them. In the following it is assumed that the range of variability of $X|\{d\}$ for $d \in \mathcal{D}$ is given as:

$$\mathcal{X}(\{d\}) = \{x_1(d), \ldots, x_N(d)\}$$

- **Constant Probability Distributions**

  A constant probability distribution $P_{X|\{d\}}$ is uniquely determined by the range of variability $\mathcal{X}(\{d\})$ of $X|\{d\}$, i. e.

$$f_{X|\{d\}}(x) = \frac{1}{|\mathcal{X}(\{d\})|} = \frac{1}{N} \tag{41}$$

- **Monotonic Probability Distribution**

Clearly, the selected probability distribution should exhibit the same monotonic behavior as the true probability distribution. The simplest exponent generating this behavior is the linear function:

$$\hat{\lambda}_0(d) + \hat{\lambda}_1(d)x$$

with

$$\begin{aligned} &\hat{\lambda}_1(d) \neq 0 \\ &\hat{\lambda}_i(d) = 0 \quad \text{for } i = 2, 3, \ldots \end{aligned}$$

where $\hat{\lambda}_0(d)$ and $\hat{\lambda}_1(d)$ have to be selected in a way that the resulting probability mass function may be used as an approximation of the true probability mass function.

Thus, the minimum amount of quantitative information which might yield a useful model consists, besides the range of variability $\mathcal{X}(\{d\})$, of the numerical values of the parameters $\hat{\lambda}_0(d)$ and $\hat{\lambda}_1(d)$. However, once $\hat{\lambda}_1(d)$ is known, the value $\hat{\lambda}_0(d)$ is given as follows:

$$\hat{\lambda}_0(d) = \log \frac{1}{\sum\limits_{i=1}^{N} e^{\hat{\lambda}_1(d)x_i}} \tag{42}$$

Thus, we have to look for a characteristic property of the true probability distribution or a key factor of the initial conditions which may be used for determining $\hat{\lambda}_1(d)$ appropriately. In terms of the deterministic variable D in its generic representation the first moment constitutes a characteristic property on the one hand and determines the value of $\hat{\lambda}_1(d)$ uniquely. Thus, the minimum amount of quantitative information in the case that the true probability distribution belongs to the monotonic family, is given by the range of variability $\mathcal{X}(\{d\})$ and the value $\mu_1(d)$ of the first moment $E[X|\{d\}]$. If $\mu_1(d)$ is known, the necessary value of $\hat{\lambda}_1(d)$ is obtained as solution of the following equation:

$$\sum_{i=1}^{N} x_i e^{\hat{\lambda}_0(d)+\hat{\lambda}_1(d)x_i} \quad = \quad \mu_1(d) \tag{43}$$

which together with (42) yields the following equation:

$$\mu_1(d) = \frac{\sum\limits_{i=1}^{N} x_i e^{\hat{\lambda}_1(d)x_i}}{\sum\limits_{i=1}^{N} e^{\hat{\lambda}_1(d)x_i}} \tag{44}$$

The minimum amount of information in terms of the deterministic variable D is given in case of a monotonic probability distribution by the value of the first moment of $X|\{d\}$, i. e.,

$$D = (E[X|\{d\}]) \tag{45}$$

- **Uni-Modal Probability Distribution**

Proceeding analogously as in the case of the monotonic family, the simplest property-conserving exponent in case of a uni-modal distribution is identified as a quadratic polynomial

$$\hat{\lambda}_0(d) + \hat{\lambda}_1(d)x + \hat{\lambda}_2(d)x^2$$

with

$$\hat{\lambda}_2(d) \neq 0$$
$$\hat{\lambda}_i(d) = 0 \quad \text{for } i = 3, 4, \ldots$$

where the normalizing constant $\hat{\lambda}_0(d)$ is uniquely determined by the values of $\hat{\lambda}_1(d)$ and $\hat{\lambda}_2(d)$.

The minimum amount of information in terms of the deterministic variable D in case of a uni-modal probability distribution is given by the values of the first and the second moment of $X|\{d\}$, i. e.,

$$D = \big(E[X|\{d\}], E[X^2|\{d\}]\big) \tag{46}$$

Equivalently, one can replace the second moment $E[X^2]$ by the variance $V[X|\{d\}]$, i. e., the second central moment of $X|\{d\}]$.

Let $\mu_1(d)$ and $\mu_2(d)$ be the actual values of the moments, and $\mathcal{X}(\{d\})$ the range of variability of $X|\{d\}$. Then, the values of the coefficients $\hat{\lambda}_i(d)$, $i = 0, 1, 2$ are obtained as solutions of the following system of equations.

$$\sum_{i=1}^{N} x_i^k e^{\hat{\lambda}_0(d)+\hat{\lambda}_1(d)x_i+\hat{\lambda}_2(d)x_i^2} = \mu_k(d) \quad \text{for } k = 0, 1, 2 \tag{47}$$

with $\mu_0 = 1$.

- **m-Extremal Probability Distribution**

  For preserving the property of the true density function of having m inner extremes by the model, a polynomial with a degree not less than $m+1$ is necessary. Thus, the polynomial which represents minimum information is given by:

$$\hat{\lambda}_0(d) + \hat{\lambda}_1(d)x + \ldots + \hat{\lambda}_{m+1}(d)x^{k+1} \tag{48}$$

  with $\hat{\lambda}_{m+1}(d) \neq 0$.

  In order to determine the coefficients $\hat{\lambda}_i(d)$ in a way that a useful approximation for the true density function is obtained, the values $\mu_i(d)$ of the first $m+1$ moments must be available. Once they are available, the coefficients $\hat{\lambda}_i(d)$, $i = 0, \ldots, m+1$, are obtained as solutions of the following system of equations:

$$\sum_{i=1}^{N} x_i^k e^{\hat{\lambda}_0(d)+\hat{\lambda}_1(d)x_i+\ldots+\hat{\lambda}_{m+1}(d)x_i^{m+1}} = \mu_k(d) \quad \text{for } k = 0, 1, 2, \ldots, m+1 \tag{49}$$

  with $\mu_0 = 1$.

The above method for selecting a probability distributions is based on the idea that the amount of necessary information depends on the degree of complexity of the given situation with respect to the probability distribution. The degree of complexity is measured by the number of extremes of the probability mass function. The larger the number of extremes the larger is the amount of necessary information.

The approach determines the minimum amount of information which is necessary to preserve the most striking properties of the true probability distribution. It is therefore named Minimum Information Principle.

Maximum Entropy Principle

Using the criterium of *minimum information* for selecting a probability distribution of one of the above given families of probability distributions is equivalent to selecting the family member with *maximum entropy.*

The Entropy of a probability distribution with finite support quantifies the amount of uncertainty represented by the considered probability distribution. It follows that a minimum information probability distribution exhibits maximum uncertainty among all other members of the considered family having the same values of the considered moments. A minimum information probability distribution belongs to the same distributional type and coincides with the most important properties of the true probability distribution. Furthermore it covers the true distribution with respect to uncertainty.

In information theory, the maximum entropy principle is proposed for selecting an unknown probability distribution:

> Consider the discrete random variable X with probability distribution P_X about which certain information are available. The maximum entropy distribution $\hat{P}_X$ is the probability distribution which is consistent with the given information and has maximum entropy, i.e., represents maximum uncertainty.

As a matter of fact, the maximum entropy principle does not include any hints about the minimum amount of information necessary to obtain a useful probability distribution and therefore using this principle is very dangerous. For example, if the true distribution is of uni-model type, but the available information is limited to the first moment, then the maximum entropy principle yields generally a monotonic distribution which is completely unusable.

Examples

1. **Throwing a Die**

   Consider a symmetric die marked on each of its six faces with a different number (from 1 to 6) of circular pits. The process consists of throwing the die and observing the number of pits on the face that is uppermost when it comes to rest. Then we have the following situation:

$$\begin{aligned} X &= \text{number of pits on the uppermost face} \\ \mathcal{X} &= \{1, 2, 3, 4, 5, 6\} \end{aligned}$$

   - Minimum Information Distribution:

     If the die is not loaded, then because of the symmetry all types of distribution can be excluded except for the family of constant distributions, implying that the minimum information distribution is given by:

$$f_X(x) = \begin{cases} \frac{1}{|\mathcal{X}|} = \frac{1}{6} & \text{for} \quad x \in \mathcal{X} \\ 0 & \text{for} \quad x \notin \mathcal{X} \end{cases} \tag{50}$$

- Maximum Entropy Distribution:

  Suppose that besides $\mathcal{X}$ also the value μ of the true value of the first moment $E[X]$ is given:

  $$\mu = 3.5$$

  Then the obtained maximum entropy distribution is the same as the minimum information distribution (50). However, if the true value of $E[X]$ is not given exactly, then the maximum entropy distribution yields a monotonic distribution, which under certain circumstances would be not usable.

2. **Monotonic Probability Distribution**

Recently a Children-Cancer Study was performed in Germany in order to answer the question, whether there is an increased probability of getting cancer for children younger than five years in the neighborhood of nuclear power plants.

Consider a nuclear power plant in Germany located in a densely populated area. Assume that the study should cover children living up to 50 km from a nuclear power plant. An unknown mechanism selects children, which subsequently get cancer. The question is whether or not the selection mechanism depends on the nuclear power plant. Of interest is the distance between the power plant and the living places of the cancer cases.

The environment of the nuclear power plant is divided into the ten disjoint areas around the power plant:

$$\begin{aligned}
K_5 &= \{x \,|\, 0 < x \leq 5\} \\
K_{10} &= \{x \,|\, 5 < x \leq 10\} \\
K_{15} &= \{x \,|\, 10 < x \leq 15\} \\
K_{20} &= \{x \,|\, 15 < x \leq 20\} \\
K_{25} &= \{x \,|\, 20 < x \leq 25\} \\
K_{30} &= \{x \,|\, 25 < x \leq 30\} \\
K_{35} &= \{x \,|\, 30 < x \leq 35\} \\
K_{40} &= \{x \,|\, 35 < x \leq 40\} \\
K_{45} &= \{x \,|\, 40 < x \leq 45\} \\
K_{50} &= \{x \,|\, 45 < x \leq 50\}
\end{aligned} \tag{51}$$

It is well-known that children may get cancer even if there is no nuclear power plant in the neighborhood. Assuming a more or less equal population density in the area considered means that with increasing i of K_i the population number.

The random variable X of interest is given as follows:

$$X = \text{ number of area code, of the next cancer case}$$

with range of variability:

$$\mathcal{X} = \{5, 10, 15, \ldots, 50\}$$

Because of the increasing population number with increasing area code, all types of probability distributions except for the monotonic family can be excluded.

- Minimum Information Distribution

  Because the true probability belongs to the monotonic family, it is sufficient to consider the first moment $E[X]$ as relevant deterministic variable:

$$D = E[X] \tag{52}$$

  with actual value μ_1. If μ_1 is known, the probability mass function of the minimum information distribution is obtained as follows:

$$f_{X|\{\mu_1\}}(x) = e^{\lambda_0(\mu_1)+\lambda_(\mu_1)x} \quad \text{for } x \in \mathcal{X}(\{\mu_1\}) \tag{53}$$

  where the range of variability is given by:

$$\mathcal{X}(\{\mu_1\}) = \{5, 10, 15, \ldots, 50\} \tag{54}$$

  and the coefficients $\lambda_0(\mu_1)$ and $\lambda_(\mu_1)$ are the unique solutions of the following system of equations:

$$\begin{aligned} \sum_{x\in\mathcal{X}(\{\mu_1\})} f_{X|\{\mu_1\}}(x) &= 1 \\ \sum_{x\in\mathcal{X}(\{\mu_1\})} x f_{X|\{\mu_1\}}(x) &= \mu_1 \end{aligned} \tag{55}$$

- Maximum Entropy Distribution

  If the same information are used as in the case of the minimum information distribution, then the same probability distribution is obtained. However, if less or more information is used, then applying the maximum entropy principle will lead to probability distribution which should not be used as it may result in a wrong decision.

3. **Uni-modal Probability Distribution**

Wind turbines for producing energy have become very popular during the last years. In order to be safe and efficient the turbines' design must stand the possible loads during its entire life.

The predicted long-term loads for wind turbines determine turbine cost and reliability and, therefore, the definition of the maximum loads, which might realistically occur during the turbine's lifetime are of considerable importance. The difficulty with predicting the extreme loads derives from the involved uncertainties. Uncertainty refers to the wind and materializes in variability in the sense that wind speed and load magnitude vary in time, even if the meteorologic conditions do not change. Therefore, the problem is to describe the variability of the loads in a way that allows reliable and at the same time accurate predictions of the maximum loads the turbines have to withstand. The predictions must have a specified reliability, as otherwise the inherent risks would be unknown and they should be accurate, as otherwise the cost of manufacturing the turbines (in case of too high load values) or the cost of occurring damages (in case of too low loads) would become excessive.

The random variable is defined as follows:

$$X_i = \text{ maximum load during one load cycle at wind condition } w_i$$

The problem is to define appropriately the wind condition w_i which represent the value of the deterministic variable (initial condition). The maximum load in a load cycle at constant wind condition varies, but excessively small values as well as excessively large values are observed only seldom. Generally, the load values concentrate in an area in between. This is a typical situation for a uni-modal probability distribution. Hence, all types of probability distributions are excluded except for the uni-modal one.

- Minimum Information Distribution

  From the Principle of Minimum Information, it is concluded that in this case an appropriate deterministic variable is given as follows:

$$D_i = (E[X_i], V[X_i]) \tag{56}$$

  Hence, for each load type of interest a finite number of wind conditions ($w_i = (\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, k$, are considered with the mean wind speed at hub height, μ_i, and the variance σ_i^2. The random variable of interest is the maximum load per cycle time on wind condition w_i denoted by X_i, where the cycle time is defined by the turbine and the load.

  In this case it seems appropriate to use a continuous approximation for X_i. Assume that the actual value (μ_i, σ_i^2) of D is given, then the variability function specifies the range of variability of X_i:

$$\mathcal{X}(\{(\mu_i, \sigma_i^2)\}) = \left\{x \,|\, \underline{x}((\mu_i, \sigma_i^2)) \leq x \leq \overline{x}((\mu_i, \sigma_i^2))\right\} \tag{57}$$

  where $\underline{x}((\mu_i, \sigma_i^2))$ and $\overline{x}((\mu_i, \sigma_i^2))$ have to be determined.

  For given (μ_i, σ_i^2) the corresponding density function is obtained as:

$$\begin{array}{l} f_{X_i|\{(\mu_i,\sigma_i^2)\}}(x) = e^{\lambda_0(\mu_i,\sigma_i^2)+\lambda_1(\mu_i,\sigma_i^2)x+\lambda_2(\mu_i,\sigma_i^2)x^2} \\ \text{for } x \in \mathcal{X}(\{(\mu_i, \sigma_i^2)\}) \end{array} \tag{58}$$

  with the coefficients $\lambda_0(\mu_i, \sigma_i^2)$, $\lambda_1(\mu_i, \sigma_i^2)$ and $\lambda_2(\mu_i, \sigma_i^2)$ are the unique solutions of the following system of equations:

$$\begin{array}{rcl} \int\limits_{x\in\mathcal{X}(\{(\mu_i,\sigma_i^2)\})} f_{X|\{(\mu_i,\sigma_i^2)\}}(x)dx & = & 1 \\ \int\limits_{x\in\mathcal{X}(\{(\mu_i,\sigma_i^2)\})} x f_{X|\{(\mu_i,\sigma_i^2)\}}(x)dx & = & \mu_i \\ \int\limits_{x\in\mathcal{X}(\{(\mu_i,\sigma_i^2)\})} (x-\mu_i)^2 f_{X|\{(\mu_i,\sigma_i^2)\}}(x)dx & = & \sigma_i^2 \end{array} \tag{59}$$

- Maximum Entropy Distribution

  Similar as in the preceding example, the maximum entropy principle yields the minimum information distribution, if the same prior information are used. However, if this is not the case, then the maximum entropy distribution may lead to a completely unusable distribution.

Bibliography

[1] Bartosyzyński and Niewiadomska-Bugaj, M. (1996): *Probability and Statistical Inference.* Wiley, New York.

[2] Bernoulli, J. (1713): *Ars conjectandi.* Basel.

[3] Bernoulli, J. (2006): *Ars conjectandi together with Letter to a Friend on Sets in Court Tennis.* Translated with an introduction and notes by Edith Dudley Sylla. The John Hopkins University Press, Baltimore.

[4] Boole, G. (1854): *The Laws of Thought.* Macmillan, London.

[5] Box, G. (1997): Scientific Method: The Generation of Knowledge and Quality. *Quality Progress*, January 1997, 47–50.

[6] Byrne, E.F. (1968): *Probability and Opinion. A Study in the Medieval Presuppositions of Post-Medieval Theories of Probability.* Martinus Nijhoff, The Hague.

[7] Cartwright, N. (1983): *How the Laws of Physics Lie.* Claredon Press, New York.

[8] Cartwright, N. (1999): *The Dappled World – A Study of the Boundaries of Science.* University Press, Cambridge.

[9] Collani, E.v. and Dräger, K. (2001): *Binomial Distribution Handbook for Applied Scientists and Engineers.* Birkhäuser, Boston.

[10] Collani, E.v. (2002): Theoretical Stochastics. In *Defining the Science of Stochastics*, Ed. E.v.Collani. Heldermann Verlag, Lemgo, 147 – 174.

[11] Collani, E.v. (2002): Empirical Stochastics. In *Defining the Science of Stochastics*, Ed. E.v.Collani. Heldermann Verlag, Lemgo, 175 – 213.

[12] Collani, E.v. and Zhai, X. (2005): *Stochastics.* (In Chinese), Beijing Publisher Group, Beijing.

[13] Collani, E.v. and Baur, K. (2007): *Was zum Teufel ist Qualität?* Heldermann Verlag, Lemgo.

[14] Davies, P. (1992): *The Mind of God: The Scientific Basis for a Rational World.* Simon & Schuster, New York.

[15] Feyerabend, P. (1988): *Against Method: Outline of an Anarchystic Theory of Knowledge.* Verso, London.

[16] Feynman, R. (1965): *The Character of Physical Laws.* MIT Press, Cambridge.

[17] Gnedenko, B.V. (1943): Les extreêmes des distributions statistiques. *Ann. Inst. H. Poincaré* 5, 115–158.

[18] Hausdorff, F. (1921): Summationsmethoden und Momentenfolgen. *Math. Z.* 9, 74-108, 280-299.

[19] Kotz S. and Johnson, N.L. (eds) (1988): *Encyclopedia of Statistical Sciences.* Wiley, New York.

[20] Kolmogorov, A.N. (1956): *Foundadtions of the Theory of Probability.* 2nd ed., translated by N. Morrison, Chelsea, New York.

[21] Levy, H. (19): *The Universe of Science.* Watts & Co., London.

[22] Pearson, K. (1900): *The Grammar of Science.* Second edition, A. and C. Black, London.

[23] Polanyi, M. (1964): *Science, Faith and Society.* The University of Chikago Press, Chikago.

[24] Schilling, P. (1991): Chance and Order in Theology. *Theology Today* 47. 1-11.

[25] Searle, J.R. (1999): The Future of Philosophy. *Phil. Trans. R. Soc.* 354, 2069 – 2080.

[26] Wulff, K. (2006): *Naturwissenschaften im Kulturvergleich. Europa – Islam – China.* Verlag Harry Deutsch, Frankfurt am Main.

Index

Abelard, Pierre, viii
accuracy, 14
aleae geometria, 33
aleatory uncertainty, 10
Altmann, Joseph, 23
ant colony, 34
Apollo 11, 15
Ars conjectandi, 15, 36, 95
astrology, 13, 19
attribute, 50
axiom, 28

belief, 19, 37
Benoulli, Jakob, 35
Bernoulli Space, 75, 87
Bernoulli Stochastics, 26
Bernoulli, Jakob, 15, 18, 83, 95
bi-variate random variable, 78
Book of Change, 13
book of Change, 84
Book of Nature, viii, 17, 18, 35
Boole, George, 37
Boorstin, Daniel J., 44
Box, George, 44
Byrne, Edmund, 24

Cartwright, Nancy, 21, 64
cause-effect chain, 7
Cayal, Ramón y, 23
complete ignorance, 99
compound random variable, 109
constant probability distribution, 107
Copenhagen interpretation, 48, 92
Cramer, Gabriel, 71

Davies, Paul, 89
determinism, 35
deterministic variable, 28, 55
deterministic world, 32, 88
Dingle, Herbert, 21
divination, 11
divine truths, 35
dynamics of evolution, 31

ecological system, 9
Eddington, Arthur S., 20
Einstein, Albert, 21, 91
electromagnetism, 33
elementary event, 76
engineering science, 23
epistemic uncertainty, 10
equi-probable, 94
Euclid of Alexandria, 28
event, 3
evolution, 3
expectation, 71

fact, 3
families of probability distributions, 107
feature, 50
Fermat, Pierre de, 32, 71
Feyerabend, Paul, vi
Feynman, Richard, 21
Feynman, Richasrd, 89
first moment, 71
Ford, Joseph, 32
freedom of belief, 19
future, 3
future aspect, 53

Galilei, Galileo, 87
gambling, 33, 94
generic deterministic variable, 111
generic representation of a deterministic variable, 111
geometry, 28
gravitational law, 21
gravity, 33
gross national product, 39
guardian of order, 24

Hamilton, William R., 86
health, 9
Hempel, Carl Gustav, v
herd instinct, 34
hexagram, 84
Holy Scripture, 17, 35
human ignorance, 58
human subjectivity, 13
Huygens, Christiaan, 32, 71

I Ching, 13

I Ging, 84
ideal gas law, 91
ignorance, 10, 40
ignorance space, 58, 75
incomplete knowledge, 98
independent, 77
initial condition, 10, 27

Jefferson,Thomas, 44

Kant, Immanuel, 24
Kaplan, Michael, 24
knowledge, 10, 31, 43
Kolmogorov, Andrey N., 76
Kuhn, Thomas S., vi

lack of knowledge, 98
law of gravitation, 90
law of nature, 7
Laws of Heredity, 31
learning process, 41
Lebesgue measure, 58
Levy, Hyman, 8, 20
logic, 26
lower quentile function, 68

mankind, 4
mathematical language, 50
mathematics, 14, 26
maximum entropy principle, 116
measurement procedure, 41, 102
Mendel, Gregor J., 31
Merzbacher, Eugen, 64
metric feature, 51
minimum information principle, 113
modern science, 8, 89
Moivre, Abraham de, 32
moment of a random variable, 70, 111
Monod, Jacques, 32
monotonic probability distribution, 108
Montmort, Pierre Rémond de, 32

neuro-science, 23
Newton Isaac, 21
Newton, Isaac, 90
nominal feature, 50

one-point prediction, 15
Oracle of Delphy, 15
ordinal feature, 51

Pareto, Vilfreda, 110
Pascal, Blaise, 32, 71
past, 3
Pearson, Karl, 5, 20
physical model, 76
Polanyi, Michael, 17, 45, 82
Popper, Karl, v
precision, 16
prediction, 11, 14
principle of order, 33
probability, 15, 35, 94
probability distribution, 105
probability distribution function, 67
probability function, 66
probability mass function, 66
probability measure, 63, 66
probability space, 76
probability survival function, 67
probability theory, 36

quantification, 48
quantum mechanics, 64
quantum physics, 23

random structure function, 63, 75
random variable, 27
randomness, 11, 32
range of variability, 53, 72
Raskic, Pasko, 24
reductionism, 8
relative frequency, 37
reliability, 14
religion, 11, 17
rule of evolution, 4
Russel, Bertrand, 20

Schilling, S. Paul, 32
Scholastic Science, viii
scholasticism, 17
Schooten, Frans, 71
Schrödinger equation, 22, 91
Schrödinger, Erwin, 22
Schrödinger's cat, 92
science, 8, 17
science of prediction, 35
scientific language, 26
Searle, John R., v
second moment, 72
σ-field, 76
simple random variables, 109
sources of uncertainty, 11
space, 4
space-time continuum, 3
speed of light, 91
spring tide, 9
state of ignorance, 60

stochastic dependence, 78
stochastic independence, 77
stochastic model, 28, 49, 75
stochastic prediction, 15
stochastic science, 55, 82, 89
strength of attraction, 63
strong interaction, 33
Sullivan, John W., 21

technology, 23
Thomson, William, 48
time, 3
trial and error, 18
truth, 7

uncertainty, 10
uncertainty space, 102
uni-modal probability distribution, 108
universal connectivity, 7, 33
unverse, 3
upper quantile function, 68

variability, 11, 30
variability function, 60, 75
variable, 50
variance, 72

weak interaction, 33
Whitehead, Alfred N., 20
Wuldd, Karl, 82

Yang, 84
Yin, 84