Journal of Uncertain Systems
Vol.2, No.3, pp.202-211, 2008
Online at: www.jus.org.uk

Defining and Modeling Uncertainty

E.v. Collani*
University of Würzburg, Sanderring 2, D-97070 Würzburg, Germany

Received January 2008; Accepted 5 March 2008

Abstract

Uncertainty about the future development constitutes the central problem of mankind. In order to solve the related difficulties many methods have been developed: divination, astrology, probability theory, statistics, possibility theory and more recently fuzzy logic, artificial intelligence and others. Each of these approaches claims to deal with uncertainty and to solve problems related to uncertainty. In this paper an alternative approach called stochastics[1], i.e., science of prediction, is introduced, which enables reliable and at the same time accurate predictions in real world situations by taking into account uncertainty about the future development.

Keywords: stochastics, prediction, reliability, accuracy, Bernoulli space

1 Introduction

Uncertainty is an ambiguous word and, therefore, it should be defined clearly at the beginning in order to avoid misunderstandings. In this paper, uncertainty is used exclusively with respect to the indeterminate future expressing the fact that in a given situation not one, but many future developments respectively events may occur.

The past is characterized by facts, which are fixed and, thus, determined. However, often facts of the past are unknown and in everyday speech this is also called uncertainty. Here, the case of unknown facts is called ignorance. Although ignorance about facts generates uncertainty about the future, it is not subsumed under uncertainty. Ignorance with respect to determinate facts and uncertainty with respect to indeterminate future events are strictly distinguished in stochastics, because they are different and, therefore, have to be handled differently. Note that in traditional science, which assumes a deterministic evolution, future is a mere transformation of past and, thus, future developments are described by simple functions of the past. Hence, past and future are equivalent and it is not necessary to distinguish between ignorance about the past and uncertainty about the future. However, if the indeterminism of the real-world evolution shall be taken into account, drawing a distinction becomes crucial.

This paper has three purposes. The first is to succinctly evaluate approaches dealing seemingly with uncertainty on the basis of their aims and methods. The second is to introduce stochastics as the science of prediction and show its potential for decision making and for solving real problems. The third purpose is a more constructive one, namely to outline the basic concepts of stochastics and introduce the stochastic model of uncertainty.

2 Science and Uncertainty

2.1 Jakob Bernoulli

About 300 years ago, Jakob Bernoulli [2,3,8] drew up the plan to inaugurate a new branch of science, which he called in Latin, “Ars conjectandi” or, in Greek, “stochastike”, which means “science of prediction”. Bernoulli had identified omnipresent “uncertainty” as a real-world aspect, causing serious problems for mankind. When he succeeded in showing that “uncertainty” could be quantified by the concept of “probability”, just as physical quantities such as weight, temperature or strength can be quantified, he was convinced that this new branch of science could and would change the world for the better. Uncertainty reveals itself by an inherent variability and, therefore, one could also talk

* Corresponding author. Email: collani@mathematik.uni-wuerzburg.de (E.v. Collani)
[1] Note that stochastics is an ambiguous term. Here it is used in its original sense of a “Science of Prediction”.

of a *science of variability*. From this point of view, the greatness of the paradigm shift proposed by Bernoulli 300 years ago becomes clear. Science had been, and still is, focussed on detecting stable - i.e. deterministic-laws, by rigorously eliminating inherent variability. Jakob Bernoulli's proposal to investigate variability implies a complete turn-around not only in science, but also in our perception of world and ways of thinking.

Unfortunately, Jakob Bernoulli passed away in 1705 before he could achieve his aims. His incomplete masterpiece was eventually published eight years after his death, in 1713, by his nephew Nikolaus Bernoulli. Subsequently, the name of this new branch of science disappeared and, with it, Bernoulli's ideas also fell into oblivion. In place of stochastics, the science of prediction, came other disciplines to seemingly fill the gap. These disciplines are probability theory and statistics and more recently for example possibility theory and fuzzy set theory.

2.2 Probability Theory

Already before the mathematician and theologian Jakob Bernoulli had turned to uncertainty, other well-known mathematicians, Pierre de Fermat and Blaise Pascal, for instance, had investigated the very special case of uncertainty which occurs in gambling. They analyzed gambling by means of an “expectation”, which represents facts, and showed that this could be done by combinatorics without the need of defining the concept of “probability”, which refers to the future development. The concept probability entered the discipline only after Jakob Bernoulli and soon the name “probability theory” became popular.

Later generations of mathematicians extended the mainly combinatorial theory by considering limits and other more sophisticated mathematical techniques. In [12], Kolmogorov notes: “The theory of probability, as a mathematical discipline, can and should be developed from axioms in exactly the same way as Geometry and Algebra”. This became a reality in 1933, when Kolmogorov himself formulated an axiomatic foundation of probability theory. Therefore, the art of probability theory is essentially that of mathematical development, starting from the axioms, and the basic concept of a *field of probabilities*. Kolmogorov notes about the meaning of the elements of a probability field the following: “What the elements of this set represent is of no importance …”. Probability theory is a special branch of mathematics aiming at further developing an abstract structure of theorems and definitions around a system of axioms. In an idealistic sense, the elements and problems dealt with have no explicit relation to anything real. The value of new results in probability theory is measured by intrinsic mathematical criteria, not by their significance for, or usefulness in application. Thus, we may conclude that probability theory does not aim at making predictions for real world processes.

2.3 Statistics

At the same time as probability theory was evolving, another discipline emerged in England, under the name “Political Arithmetic”. This discipline aimed at collecting and analyzing data, e.g., about “Value of Lands, People, Buildings: Husbandry, Manufacture, Commerce, Fishery, Artizans, Seamen, Soldiers; Publick Revenues, Interest, Taxes, Superlucration, Registries, Banks Valuation of Men, Increasing of Seamen, of Militia's, Harbours, Situation, Shipping, Power at Sea, &c.'” as William Petty [13], one of the founders of political arithmetic, notes.

At the end of the 18th century political arithmetic became better known as “statistics”, which is an artificial word, created in Germany and used as a name for a school devoted to the academic study of state constitutions. In contrast to probability theory, the discipline “Statistics” remained an applied one. It was used to extract information from large data sets, to support the decision-making process. One of the main findings of statistics, when dealing with large sets of data, was a certain stability of averages. In early times, this stability was thought to represent “divine” laws. As a matter of fact, the discovery of such stable laws became a central aim of statistics, which is achieved by reducing the observed variability.

In contrast to probability theory, concepts and procedures in statistics are introduced and developed for application. However, the lack of a unified theory does not allow the adoption of applicability and usefulness as evaluation criteria. Even the basic notion of “probability” is not consistently defined, a fact that has led to fierce controversies within the statistical community. There are different and opposing schools of thought in statistics which, like religions, appear to be based on dogmata and beliefs. The entire history of statistics is characterised by continuous controversies about the meaning of concepts and about the relevance of evaluation criteria.

2.4 Other Attempts

More recently new approaches addressing “uncertainty” were proposed, among them fuzzy set theory, possibility theory and the artificial intelligence approach. These new disciplines have in common that the word “uncertainty” is

mainly used with respect to unknown facts. Consequently, the central aim of these approaches is directed to the past and not to the future, similar as in the case of statistics.

The orientation to the past implies that subjective knowledge about the past plays a central role. Thus, the degree of belief or similar concepts are modeled and used for deriving statements. Clearly, these approaches cannot lead to a science of prediction.

2.5 Stochastics

Most investigative activities can be classified into two broad categories on the basis of their corresponding aims. One type of activity aims at disclosing a determinate, but unknown fact and, thus, reducing ignorance, whereas the other type aims at predicting or manipulating an indeterminate future development and, thus, reducing uncertainty. Although rather different, both categories are either explicitly or implicitly reliant on predictions, implying that the capacity to make reliable predictions is decisive for success. It was probably this perception that prompted Jakob Bernoulli to tackle the problem of developing a scientific approach to prediction.

The need for reliable and at the same time accurate predictions cannot be satisfied by probability theory and statistics on the one hand, or by the more recent approaches on the other. Probability theory develops “mathematical concepts” independently of their usefulness. Statistics develops methods for analyzing large data sets in order to detect stabilities. Finally, the new approaches center rather on the past than on the future. In contrast, stochastics represents a conceptual and theoretical basis covering all aspects which are involved in the scientific process of making predictions about future developments. Stochastics aims at quantifying and investigating uncertainty, i.e. variability, in order to make reliable and accurate predictions in real-world situations. The ultimate aim is to arrive at better decisions and, therefore, the usefulness of “models” and procedures is of central importance.

3 Quantification of Uncertainty

There are many branches of “Science”, which are quantified and which are commonly referred to as “exact” or “mathematical” sciences. Nominally at least, each of these branches deals with its own specialist part of the real world and sets out to establish formal relationships between various quantities and elements within its area of ownership. Of course, the parts overlap one another to a certain extent, and the findings of one branch can be of use in other branches. Comparing stochastics with these more traditional branches of science reveals two major differences.

1) Although regarded as unique and unified areas, nonetheless, the traditional branches are frequently associated with a clear partitioning of the activity within them, along such lines as “macro/micro”, “fluids/solids”, etc., each with clearly different aspects. In contrast, stochastics is fully defined by a single aspect, namely “uncertainty”.

2) While the aspects investigated by other branches overlap to a certain extent, it is generally possible to assign different aspects to different sciences. The aspect “uncertainty”, however, is omnipresent; and every branch of science should have to cope with it.

Whilst uncertainty is a single aspect, it occurs in almost infinitely many forms. Moreover, humans have no tangible sense of uncertainty. It cannot be seen, tasted, heard, smelled or felt; it can be experienced only indirectly. Therefore, identification of the relevant quantities and elements is much more difficult and quantification becomes a major task. These difficulties aside, the problem has a positive feature. The fact that the subject of stochastics pertains to a single and universal aspect of the real world, enables the development of a unified and general (mathematical) model for quantifying it. To develop this model the relevant variables must first be identified, and then functions have to be derived to represent the pertinent dependencies and relations.

3.1 Variables

Those quantities and elements of the process which are of interest have to be identified and represented by variables, i.e., by symbols and appropriate ranges.

3.1.1 Random Variables

Let us first consider the terminal point of the process and let X denote the variable describing the quantity of interest with range of variability $\mathcal{X}$. Note that the range of variability $\mathcal{X}$ is necessarily bounded, because this is a characteristic property of any real-world quantity.

In the terminal point of the process, the variable X will adopt one element of $\mathcal{X}$. However, the value to be realized is subject to the "randomness" arising from indeterminate future developments. Adopting this terminology, it is appropriate that X be termed a *random variable*.

As to the range of variability $\mathcal{X}$ of X, it contains any observable value which might occur in the given situation. As a consequence the set $\mathcal{X}$ is necessarily finite. However, if possible the discrete set $\mathcal{X}$ is often replaced by a continuous and bounded set.

3.1.2 Deterministic Variables

Typically, any future development is related to the starting point. Thus, those aspects of the starting point which have an impact on the future development of interest, and, hence, on X, have to be specified. Let D be the variable describing the relevant quantities in the starting point, i.e., the initial conditions, and let $\mathcal{D}$ denote the set of admitted values of D. In a given situation, the initial conditions $d_0 \in \mathcal{D}$ are fixed, i.e., determinate by the past development. Therefore, D is referred to as a *deterministic variable*.

The pair (X,D) defines the necessary variables. The random variable X represents the aspects of interest in the future and the deterministic variable D the aspects of the past, which are relevant for X. The identification of X and D is of utmost importance for successfully quantifying uncertainty.

3.1.3 Example for Selecting the Variables

Consider a production process, which produces items independently and under identical conditions. It is known that most of the items produced will be conforming to given specifications, but some of them might be nonconforming. Of interest is the number of nonconforming items, which will be produced in the considered production run. In this case the random variable X is readily obtained:

X = Number of nonconforming item.

The deterministic variable D determines completely the random performance of X, which is given by the corresponding probability distribution. Thus, the following deterministic variable is obtained:

$D = (D_1,D_2)$ with

D_1= Size of the production run.

D_2= Probability of producing a nonconforming item.

3.2 Functions

In all realistic cases, future development is related to the initial conditions, and this relationship must be quantified. The initial conditions, represented by the value d of the deterministic variable D, determine the range of variability $\mathcal{X}$ of the random variable X and the probabilities of all events with respect to X. An event E with respect to X, is a subset of $\mathcal{X}$. The probability of E is a real number between 0 and 1, which expresses the degree of certainty that the event E will occur. Just as the actual value of the temperature of an object is determined by some initial conditions, so too, by analogy, will the probability of an event be fixed by the initial conditions.

3.2.1 Variability Function

To distinguish between different initial conditions, the random variable X under the condition d is introduced and denoted by $X|\{d\}$ with range of variability $\mathcal{X}(\{\mathrm{d}\})$. Thus, a first function (cf [6]) denoted $\mathcal{X}$ and called *variability function* is identified, which assigns to each considered situation the corresponding range of variability.

3.2.2 Example for a Variability Function

In the above example the initial conditions d are given by the size n of the production run and the nonconformance probability p. Any value of $d = (n, p)$ determines the range of variability of the random variable $X|\{d\}$ and, thus, the corresponding variability function $\mathcal{X}$:

$$\mathcal{X}\left(\{\mathrm{d}\}\right) = \mathcal{X}\left(\{(\mathrm{n},\mathrm{p})\}\right) = \left\{\mathrm{x} \in \mathbb{N} \,\middle|\, 0 \le \mathrm{x} \le \mathrm{n}\right\} \qquad \text{for any p}.$$

3.2.3 Random Structure Function

Besides the extent of variability of X the structure of randomness described by the probabilities of events is of decisive significance. Thus, the function $\mathcal{P}$ called *random structure function* is introduced, which assigns to each possible situation with respect to the admitted values d of D the corresponding probability distribution of $X|\{d\}$:

$$\mathcal{P}\left(\{\mathrm{d}\}\right) = P_{\mathrm{X}|\{\mathrm{d}\}} \tag{1}$$

Note that the quantification above constitutes a far-reaching extension of the classical approach for instance in physics. Traditional science assumes a deterministic, i.e. a one-to-one, relationship between initial conditions and the outcome of a process and the usually observed deviations between calculated and observed results are typically ascribed to measurement error, thereby obviating the need for any consideration of (process) randomness. In stochastics, the deterministic relations of classical physics are replaced by stochastic relations that explicitly force an investigation of the observed variability.

3.2.4 Example for a Random Structure Function

Again the example of the production run is considered. The random structure function $\mathcal{P}$ assigns to any initial condition given by the actual value $d = (n,p)$ of the deterministic variable D the corresponding probability distribution $\mathrm{P}_{X|\{d\}}$ of the random variable $X|\{d\}$. The random structure function in the example is given by:

$$\mathcal{P}\left(\{\mathrm{d}\}\right) = \mathcal{P}\left(\{(\mathrm{n},\mathrm{p})\}\right) = P_{\mathrm{X}|\{(\mathrm{n},\mathrm{p})\}} \qquad \text{with} \qquad \mathrm{P}_{\mathrm{X}|\{(\mathrm{n},\mathrm{p})\}}\left(\{\mathrm{x}\}\right) = \binom{\mathrm{n}}{\mathrm{p}} \mathrm{p}^{\mathrm{x}} \left(1-\mathrm{p}\right)^{\mathrm{n-x}} \qquad \text{for } \mathrm{x} \in \mathcal{X}\left(\{(\mathrm{n},\mathrm{p})\}\right).$$

3.3 Generic Representation of the Deterministic Variable

As mentioned above, for any realistic situation the range of variability $\mathcal{X}\left(\{d\}\right)$ of $X|\{d\}$ is bounded and strictly speaking even finite. Therefore, in the univariate case[2], the probability distribution $\mathrm{P}_{\mathrm{X}|\{d\}}$ of $X|\{d\}$ is uniquely determined[3] by the sequence

$$(\mu_1(d), \mu_2(d), \cdots) \tag{2}$$

of values of the moments $E\left[X^k \,|\, \{d\}\right]$, $k = 1,2,\ldots$. 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:

$$\mathrm{D} = \left(\mathrm{E}[X], \mathrm{E}\left[X^2\right], \mathrm{E}\left[X^3\right], \ldots\right) \tag{3}$$

with actual value $d = (\mu_1, \mu_2, \mu_3, \ldots)$.

The generic representation (3) of D represents not single factors having an impact on the future development with respect to X, but the entire past with all its known and unknown dimensions of relevance.

3.4 Generic Representation of Probability Distributions

With (3) an appropriate representation of the deterministic variable D has been identified. Next, a generic representation of the corresponding probability distribution must be selected.

The probability distribution $P_{X|\{d\}}$ is determined by the relevant initial conditions given by the value $d = (\mu_1, \mu_2, \ldots)$ of the deterministic variable $D = \left(E[X], E\left[X^2\right] \ldots\right)$ and the corresponding range of variability

[2] In the multivariate case an analogous statement holds.

[3] For a proof see Hausdorff [10].

$\mathcal{X}(\{d\})$ of the random variable $X|\{d\}$. It is easy to show that any probability mass function of a probability distribution $P_{X|\{d\}}$ can be represented in the following form:

$$f_{X|\{d\}}(x) = e^{\lambda_0(d)+\lambda_1(d)+\lambda_2(d)+\cdots} \tag{4}$$

where generally only a finite number of coefficients $\lambda_i(d)$ are unequal to zero.

The coefficients $\lambda_i(d)$, $i = 0,1,2,\ldots$, are obtained[4] as the solution to the following system of equations:

$$\mu_i = \int_{\mathcal{X}(\{\mathrm{d}\})} \mathrm{x}^i e^{\sum_{k=0}^{\infty}\lambda_k(\mathrm{d})x^k} \, \mathrm{dx} \tag{5}$$

with $\mu_0 = 1$.

The generic representation (4) of $f_{X|\{d\}}$ holds for all probability distributions describing real processes. Note that for initial conditions given in the generic representation (3), the determination of the probability distribution by means of (5) is straightforward, although not simple.

If the polynomial in the exponent of (5) is of degree 1, then the value μ_1 of the first moment together with $\mathcal{X}(\{d\})$ determines completely the probability distribution $P_{X|\{d\}}$. Similar, if the polynomial is of degree 2, then $\mathcal{X}(\{d\})$ and the values μ_1 and μ_2 are necessary and sufficient for determining $P_{X|\{d\}}$.

3.5 Classification of Probability Distributions

The here developed classification scheme for probability distributions $P_{X|\{d\}}$ is based on situations which lead to different shapes of the probability mass function $f_{X|\{d\}}$. We distinguish between simple and compound processes and the corresponding random variables. A compound process consists of at least two steps, where the process of the first step determines which of at least two different processes and hence random variables of the second step is run, and so on. In contrast, a simple process consists of only one process, which runs always in the same manner.

In case of a simple process, it is sufficient to distinguish three cases with respect to the probability distribution:

- **Case 1:** The first one is characterized by the fact that each of the possible outcomes has the same probability of occurrence yielding a constant probability mass function. In this case the polynomial in the exponent of (4) is given by a constant.
- **Case 2:** The second case refers to processes, where the probability of occurrence increases (decreases) with an increasing value of the possible outcome. The resulting probability mass function has one boundary maximum. In this case the polynomial in the exponent of (4) increases (decreases) for $\mathrm{x} \in \mathcal{X}(\{\mathrm{d}\})$.
- **Case 3:** Finally, the third case is characterized by the fact that the extreme values of X have small probabilities and values somewhere in between have maximum probabilities. The resulting probability mass function has exactly one inner maximum. In this case the derivative of the polynomial in the exponent of (4) has exactly one zero for an inner point of $\mathcal{X}(\{d\})$ with a negative second derivative.

Compound processes combine stepwise simple processes yielding probability mass functions with possibly several maxima. For instance combining a monotonic increasing and a monotonic decreasing density function generally yields a probability mass function of bathtub shape with two boundary maxima. Combining two uni-modal distribution generally results in a probability mass function with two inner maxima. In this paper only the case of simple processes shall be dealt with in detail. The three possible distribution functions are analyzed below.

3.5.1 Family of Constant Probability Distributions

The simplest case is characterized by the fact that there are no points of inclination or disinclination of $f_{X|\{d\}}$ implying that the probability mass function is constant.

[4] See Hausdorff [10].

In this case the range of variability $\mathcal{X}(\{d\}) = \{x_1(d), x_2(d), \ldots, x_{N(d)}(d)\}$ with $x_i(d) < x_{i+1}(d)$ for $i = 1, \ldots, N(d) - 1$ constitutes the only quantitative information necessary for determining "exactly" the probability distribution:

$$f_{X|\{d\}}(x) = \begin{cases} \dfrac{1}{|\mathcal{X}(\{d\})|}, & \text{for } x \in \mathcal{X}(\{d\}) \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

In view of the representation (4) the following values of the distribution parameters are obtained:

$$\begin{aligned} \lambda_0(d) &= -\log|\mathcal{X}(\{d\})| \\ \lambda_k(d) &= 0 \quad \text{for } k = 1, 2, \ldots \end{aligned} \tag{7}$$

The conditions (7) on the values of $\lambda_i(d)$ define the *family of constant probability distributions*.

Consider the continuous approximation of a constant probability distribution with range of variability $\mathcal{X}(\{d\}) = \{x \in \mathbb{R} \mid a(d) \le x \le b(d)\}$. Then the actual value of the deterministic variable *D* in its generic representation is given by:

$$d = (\mu_1, \mu_2, \cdots) \text{ with } \mu_k = \frac{1}{k+1} \frac{b(d)^{k+1} - a(d)^{k+1}}{b(d) - a(d)} \quad \text{for } k = 1, 2, \cdots. \tag{8}$$

Clearly in the case of a constant probability distribution the range of variability constitutes the necessary and sufficient information. For obtaining the range of variability the actual values of the first and second moment are needed.

3.5.2 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 probability mass functions. There are two cases "the smaller the more probable", and, "the larger the more probable". In the former, the exponent of the probability mass function given by (4) decreases monotonically, whilst in the latter it increases monotonically.

Thus, the *family of monotonic distributions* is defined by the following condition for $x \in \mathcal{X}(\{d\})$:

$$\sum_{i=1}^{\infty} i\lambda_i(d)x^{i-1} \begin{cases} < 0, & \text{monotonic decreasing} \\ > 0, & \text{monotonic increasing.} \end{cases} \tag{9}$$

3.5.3 Family of Uni-Modal Probability Distributions

The next level of complexity is reached, if the probability mass function has exactly one maximum in the interior of the range of variability. 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 family of uni-modal probability distributions is formally defined by the following conditions on the parameter values λ_i: There is exactly one $x_1 \in \mathcal{X}(\{d\})$ with

$$\sum_{i=1}^{\infty} i\lambda_i(d)x_1^{i-1} = 0, \sum_{i=2}^{\infty} i(i-1)\lambda_i(d)x_1^{i-2} < 0. \tag{10}$$

The variability of a very large number of random variables follows a uni-modal probability distribution. Moreover, under rather general conditions the probability distribution of a sum of arbitrary random variables is uni-modal. Very often this property is obtained already for a rather small number of components. Thus, one can state that the uni-modal probability distributions have paramount importance.

4 Stochastic Modelling

The above described quantification of uncertainty is the first step for stochastic modeling. The second step is to derive a stochastic model, which describes as well as possible the existing uncertainty based on the available information.

As a prelude to tackling issues relating to variables, functions, etc., we note again that the source of human uncertainty may be divided into two broad classes: "randomness" and "ignorance". This is best appreciated by focussing on the central theme of stochastics - the science of prediction. Consider a phenomenon or process which has a starting point and a point of termination. The actual state in the starting point determines the initial conditions for the process. These initial conditions are fixed by past events. The process state in the terminal point - for which a prediction is sought - depends on these initial conditions. It also depends on future developments, which are, as yet, indeterminate. Clearly then, the two elements producing uncertainty and impacting on the accuracy of the prediction are, firstly, any lack of knowledge ("ignorance") concerning the relevant past represented by *D* and, secondly, the indeterminate nature of future developments ("randomness") represented by *X*. This division of uncertainty is central to an understanding of the modeling approach adopted in stochastics.

A full representation, or model, of a real-world (stochastic) situation must be built around the pair of variables (*X, D*). It involves:

- The ignorance concerning the actual value of *D*. The ignorance is described by a set $\mathcal{D}$, which consists of all those values of *D*, which cannot be excluded based on the available experience about the given situation. The set $\mathcal{D}$ is called *ignorance space*, where each of its elements stands for different initial conditions.
- The corresponding ranges of variability $\mathcal{X}(\{d\})$ for $X|\{d\}$, with $d \in \mathcal{D}$, given by the variability function $\mathcal{X}$. The variability function $\mathcal{X}$ is determined similarly as the ignorance space. Each image set $\mathcal{X}(\{d\})$ consists of those values for $X|\{d\}$, which cannot be excluded in the situation given by *d*.
- The stochastic relations between *D* and *X* are specified by the random structure function $\mathcal{P}$, which assigns to each initial condition *d* the corresponding probability distribution $\mathcal{P}(\{d\}) = P_{X|\{d\}}$. In order to allow reliable and accurate predictions the images $P_{X|\{d\}}$ must reflect sufficiently well the actually existing uncertainty.

The stochastic model denoted $\mathbb{B}_{X,D}$ is characterized by three components:

$$\mathbb{B}_{X,D} = (\mathcal{D}, \mathcal{X}, \mathcal{P}) \ . \tag{11}$$

In commemoration of Jakob Bernoulli, the *hyperspace* defined by this set is labelled as *Bernoulli-Space*. The Bernoulli-Space contains those real-world quantities which are necessary for describing uncertainty in a given situation from the stochastic point of view. It is worth noting that since the degree of ignorance is explicitly factored into the model, via $\mathcal{D}$, the model may be used recursively as new knowledge becomes available.

The first component $\mathcal{D}$, i.e., the *ignorance space* reflects the existing level of ignorance in a given situation. The *variability function* $\mathcal{X}$ specifies the extent of variability of the future development, while the *random structure function* $\mathcal{P}$ defines the random structure on the range of variability. Next, methods for determining in particular the *random structure function* have to be developed.

5 Random Structure Function

As stated in the previous section, the moments of a random variable *X* together with the range of variability determine uniquely the probability distribution $P_{X|\{d\}}$ and, therefore, the moments may be selected as deterministic variable *D*. Unfortunately, in general the values of only a few moments can be known. Therefore, a criterion must be developed for selecting an appropriate random structure function $\mathcal{P}$ assigning to the truncated value of the deterministic variable a probability distribution, which can be used as an appropriate substitute of the true one. Probability distributions are basically used for making predictions. Consequently, an approximation is said to be appropriate, if it allows reliable and accurate predictions.

Clearly the amount of information necessary for obtaining an appropriate random structure function depends on the complexity of the situation. If the situation is simple, the necessary amount of information is small, conversely in a complex situation, the necessary amount is larger.

For the sake of simplicity the argumentation is limited here to the univariate case and for probability distributions, which might be constant, monotone or uni-modal.

The main subject of investigation in 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. As shown above these patterns are expressed essentially by the number and type of extremes of the exponent of (4). Number and type of extremes refer to points of inclination and points of disinclination of $X|\{d\}$ and represent common experiences with processes of interest. 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 values of the extremes.

5.1 Minimum Information Distribution

In order to check the usefulness of any approximation its properties must be compared with those of the true probability distribution.

A probability distribution describes the uncertainty about the future development by stating the extent of variability and the corresponding random structure. Both determine the magnitude of uncertainty.

An appropriate approximation should have the same relevant features as the true probability distribution and the size of uncertainty of the approximation should be at least as large as that of the true probability distribution. The second requirement is necessary, because a smaller size of uncertainty of the approximation would be tantamount of having used wrong assumptions. The size of uncertainty of a probability distribution is measured by its entropy. Thus, it is required that the entropy of the approximation is at least as large as that of the true probability distribution.

The relevant features of a probability distribution are given by its location and its shape. The location of the probability distribution of X is specified by its range of variability $\mathcal{X}$, while the shape is determined by the polynomial in the exponent of (4). Thus the required properties of an approximation are conservation of location and shape, which are given by the range of variability and the number and types of maxima of the defining polynomial in the exponent of (4). Evidently, a polynomial of degree 0 is necessary for the constant case, a polynomial of degree 1 is necessary in the monotone case, and a polynomial of degree 2 in the uni-modal case. In case of a constant probability distribution the required information yields the exact probability distribution, while in the two remaining case approximations are obtained.

For the determination of an appropriate polynomial of degree 1 in the monotone case the value $\mu_1(d)$ is necessary, and for the determination of an appropriate polynomial of degree 2 in the uni-modal case the values of $\mu_1(d)$ and $\mu_2(d)$. Thus, the following sets of necessary information are obtained in terms of the generic deterministic variable:

Table 1: Minimum information for obtaining appropriate approximations

Type of $f_{X\|\{d\}}$	minimum information
constant	$\mathcal{X}(\{d\})$
monotone	$\mathcal{X}(\{d\}),\ \mu_1(d)$
uni-modal	$\mathcal{X}(\{d\}),\ \mu_1(d),\ \mu_2(d)$

The approximations obtained based on the minimum information are called *minimum information distributions.* The requirement that $f_{X|\{d\}}$ agrees with the essential properties of the true probability distribution specified in Table 1 guarantees a good approximation with respect to the random structure. However, the question arises whether or not the approximation underestimates the uncertainty represented by the true, but unknown probability distribution. Note that a probability distribution quantifies the structure of randomness, while the amount or size of the generated uncertainty is quantified by its entropy. The entropy of a probability distribution is zero for a degenerated one-point probability distribution and maximum for the uniform or constant probability distribution. It was introduced by Claude Shannon in 1948 by stating clear requirements for a measure of the size of uncertainty and showing that there is essentially only one type of function meeting theses requirements. He called this function *entropy* as it coincides with the famous Boltzmann entropy in physics, which is a measure of disorder. Thus, the size of uncertainty of a probability distribution $P_{X|\{d\}}$ is measured by its entropy $H(P_{X|\{d\}})$ given by

$$H\left(P_{X|\{d\}}\right) = -\sum_{x\in\mathcal{X}(\{d\})} P_{X|\{d\}}\left(\{x\}\right)\log P_{X|\{d\}}\left(\{x\}\right). \tag{12}$$

In order not to underestimate the existing uncertainty, the approximation should represent at least the same magnitude of uncertainty as the true probability distribution. It is easy to show that the above defined *minimum information distributions* have maximum entropy among all distributions of $X|\{d\}$ with the range of variability

$\mathcal{X}(\{d\})$ and the corresponding values of the moments. Therefore, the random structure function is easily obtained by fixing the range of variability and the values of the corresponding moments.

6 Improvement of the Stochastic Model

Any Bernoulli-Space represents the momentarily available information or equivalently the current ignorance. If the current information is small, the model gives only a very vague description of reality, which means that the predictions derived from the model, although reliable, are often too inaccurate and therefore useless. In such cases the stochastic model must be improved, where the improvement refers particularly to the ignorance space $\mathcal{D}$ and the variability function $\mathcal{X}$.

Model improvements are performed by means of stochastic measurement procedures. In [7] measurement procedures for improving the ignorance space are developed, while in a series of papers [14,15,16,17,18] the problems of improving the variability function are investigated.

7 Summary

In order to model appropriately uncertainty about future developments, it is necessary to include the two sources of uncertainty namely ignorance as being a characteristic property of man and randomness being an inherent property of universe. Ignorance refers to the fixed past, while randomness describes the indeterminism of the future.

The stochastic model, i.e., the Bernoulli-Space, includes ignorance and randomness and is based only on available experience and gets by without unverified assumptions. As a consequence the model covers the existing uncertainty, which is a necessary condition for reliable predictions.

There are two classes of stochastic procedures, the first one yielding statements about the indeterminate future, while the second one yields statements about the determinate past. The former is called *prediction class* the latter is called *measurement class*. It is a characteristic feature of stochastics that both classes of procedures are based on predictions enabling the development of a unified theory for modelling and handling human uncertainty.

References

[1] Bernoulli, J., *Ars Conjectandi Basileae*, 1713.

[2] Bernoulli, J., Wahrscheinlichkeitsrechnung (Ars conjectandi), übersetzt von R. Haussner, Ostwalds Klassiker der exakten Wissenschaften 107 u. 108, Leipzig: Akadem. Verlagsanstalt, 1899.

[3] Bernoulli, J., Die Werke von Jakob Bernoulli, Bd. 3. Bearbeitet von B.L. van der Waerden, Basel: Birkhäuser, 1975.

[4] v.Collani, E., History, State of the Art and Future of Stochastics, In: Ivor Grattan-Guinness, B.S. Yadav (eds): History of the Mathematical Sciences, New Delhi: Hindustan Book Agency, pp.171-194, 2004.

[5] v.Collani, E., and K. Dräger, *Binomial Distribution Handbook for Scientists and Engineers*, Birkhäuser, Boston, 2001.

[6] v.Collani, E., Theoretical Stochastics, In: v. Collani, E. (ed.): Defining the Science of Stochastics, Heldermann Verlag, Lemgo, pp.147-174, 2004.

[7] v.Collani, E., Empirical Stochastics, In: v. Collani, E. (ed.): Defining the Science of Stochastics, Heldermann Verlag, Lemgo, pp.175-213, 2004.

[8] v.Collani, E., Jakob Bernoulli deciphered, *Bernoulli News*, vol.13, no.2, pp.2-9, 2006.

[9] Guiasu, S., *Information Theory with Applications*, McGraw Hill, New York, 1977.

[10] Hausdorff, F., Summationsmethoden und Momentenfolgen, *Math. Z.*, vol.9, pp.74-108, pp.280-299, 1921.

[11] Kolmogorov, A.A., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Julius Springer, Berlin, 1933.

[12] Kolmogorov, A.N., *Foundations of the Theory of Probability*, Translation B N. Morrison, Chelsea Publishing Company, New York, 1956.

[13] Petty, W., *Political Arithmetick*, London, 1690.

[14] Sans, W., X. Zhai, *et al.*, Improving the Variability Function in Case of a Monotonic Probability Distribution, *Economic Quality Control*, vol.20, pp.121-142, 2005.

[15] Sans, W. and E.v. Collani, A Note on Monotonic Distributions, *Economic Quality Control*, vol.18, pp. 67-81, 2003.

[16] Sans, W., D. Stübner, *et al.*, Improving the Variability Function in Case of a Uniform Distribution, *Iapqr Transactions*, vol.30, pp.1-12, 2005.

[17] Schelz, F., W. Sans, and E.v. Collani, Improving the Variability Function in Case of a Uni-Modal Probability Distribution, *Economic Quality Control*, vol.22, pp.19-39, 2007.

[18] Stübner, D., W. Sans, *et al.*, Measurement Procedures for Improving the Variability Function, *Economic Quality Control*, vol.19, pp.215-228, 2004.