A Model for *n*-dimensional Bayes Nets*

Un modelo para redes bayesianas *n*-dimensionales

Wolfgang Balzer†
Klaus Manhart‡

Abstract

We reconstruct the core theory for n-dimensional Bayes nets in a clear and complete way in the structuralist framework. Our formulation also includes the n one-dimensional probability spaces and one n-dimensional probability space. We explain the intended systems and the empirical claim of the theory of n-dimensional Bayes nets. In our formulation, the concept of dimension is highlighted, and we discuss whether events or sentences (statements, propositions) could be used as the basic building blocks.

Keywords: Bayes nets - structuralism - dimension

Resumen

Reconstruimos la teoría central de las redes bayesianas *n*-dimensionales de forma clara y completa en el marco estructuralista. Nuestra formulación también incluye los n espacios de probabilidad unidimensionales y un espacio de probabilidad n-dimensional. Explicamos los sistemas previstos y la afirmación empírica de la teoría de las redes bayesianas n-dimensionales. En nuestra formulación, se destaca el concepto de dimensión, y discutimos si los eventos o las oraciones (enunciados, proposiciones) podrían usarse como bloques de construcción básicos.

Palabras clave: redes bayesianas - estructuralismo - dimensión

* Received: 3 July 2021. Accepted in revised version: 13 September 2021.
† MCMP, LMU München. To contact the author, please write to: balzer@lrz.uni-muenchen.de.
‡ To contact the author, please write to: mail@klaus-manhart.de.
Metatheoria 12(1)(2021); 1-18. ISSN 1853-2322. eISSN 1853-2330.

1. Introduction

The notion of probability was introduced in the 17th century in studies of card playing, and in the 18th century, the first statistics were developed for handling socio-political problems, for instance for the health sector. In the beginning of the last century, the theory of probability was formulated in a way that is used today (see Kolmogorov 1950). Some years later, the notion of *conditional probability* was introduced and investigated, for instance by Renyi (1970). This notion became important for physics, especially for statistical mechanics. Here, *Markov processes* (Bremaud 1999) were used to describe linear processes in time. In the middle of the last century, when nets came into focus, a new kind of process came to the stage. These processes can be divided into several branches and were investigated and described probabilistically (Cox 1947).

In 1763, a theorem about probabilities was published by T. Bayes, which became important at the end of the last century in expert systems (Michie 1979). This led to the notion of a Bayes net, which was introduced in 1985 by J. Pearl (see, for instance, Pearl 1988). In a Bayes net, the theorem of Bayes plays a central role in the deduction and calculation of probabilities in a net. This role spread to statistics, computer science, and simulation (Sokolowski & Banks 2010). Today the rather simple theorem of Bayes is common knowledge. Bayes nets are applied in all scientific domains: in psychology, decision theory, AI, robotics, and computer nets and on the internet. They are also used in many examples: learning, the confirmation of facts and beliefs, the simplification of beliefs, or the building or changing of beliefs.

In a Bayes net, lists of probabilities (lists of numbers) are studied. In such a list, one number can depend on another number or numbers. Time can also be considered. A Bayes net visualizes the dependency structure among variables under investigation. The variables are represented by nodes and dependencies by connections among the nodes. Missing connections imply independence among the variables.

One can say that in a net, a probability that is attached to a node can be recalculated using conditional probabilities. In this way, in a net, probabilities are changed, introduced, eliminated, and also determined. These different kinds of changes in a net can be expressed in a graphical mode as well as by lists of numbers. This diversity is a central topic in Bayes nets. Exactly at this point, new aspects emerge that were not seen before in probability theory.

The formulation of the hypotheses for a Bayes net does highlight the structure of the net but "outsources" the details of using probability spaces. In the literature about Bayes nets, the notion of an n-dimensional probability space is used but rarely explained. Probability spaces are just presupposed. In our reconstruction, the conceptual apparatus of probability theory is also included.

Our reconstruction uses the structuralist framework, which is found in the theory of science and which is described in many works (see for instance Balzer, Moulines & Sneed 1987 or Abreu, Lorenzano & Moulines 2013). We assume that a scientific theory can be described in an idealized way by a set $\mathbf{M}$ of models, a set $\mathbf{M}_p$ of potential models, a set $\mathbf{M}_{pp}$ of partial models, and a set $\mathbf{I}$ of intended systems. A potential model x consists of a list of base sets $B_1,\ldots, B_m$, a list of auxiliary sets $C_1,\ldots, C_l$, and a list of relations $R_1,\ldots, R_n$, which are typified by $\tau_1,\ldots, \tau_n$:

$$x = \langle B_1,\ldots, B_m, C_1,\ldots, C_l, R_1,\ldots, R_n\rangle.$$

Some of the potential models are also models of the theory. Partial models y are systems that can be embedded in potential models x; y is a restriction of components and of the appertaining numbers:

$$y = \langle B^*_1,\ldots, B^*_{m^*}, C^*_1,\ldots, C^*_{l^*}, R^*_1,\ldots, R^*_{n^*}\rangle \sqsubseteq \langle B_1,\ldots, B_m, C_1,\ldots, C_l, R_1,\ldots, R_n\rangle = x,$$

where $m^* \leq m$, $l^* \leq l$ and $n^* \leq n$, and $B^*_1 \subseteq B_1,\ldots, R^*_{n^*} \subseteq R_n$.

An intended system for a theory is a real system, which is scientifically investigated by a group of researchers. The intended system contains several scientifically determined observations (facts) that are described by notions used for the given theory.

2. The function of a Bayes net

A Bayes net consists of n nodes, directed lines, random variables, and an n-dimensional probability space. To a node, exactly one random variable is assigned. As the lines have a unique direction, they can be represented graphically by arrows.

A random variable X can be expressed in two ways. It can run through a set of events or through a set of sentences (or statements or propositions). X is a function that assigns to each event (or to each sentence) a function value. These values represent probabilities — numbers between 0 and 1. In simple cases, all possible function values of a random variable X for a node are described in matrix form and can be attached to the node to which the random variable belongs. A Bayes net contains, therefore, n sets $V_1,\ldots, V_n$ of values, where the values belong to events (or sentences). From the n sets of values, the Cartesian product is built. A list of values from this Cartesian product is called a state of the Bayes net. A state consists, in other words, of n components, where each component is a value belonging to an event (or a sentence).

An arrow begins at one node and ends at another node. If a state z of the net and an arrow from node i to j are given, a connection from i to j can be described as follows. We look at the i-th component of z, the value v_i, and at the j-th component of z, the value v_j. By the machinery of the Bayes net, the value v_j can be changed depending on the value vi. A new value v_{j^*} for the j-th component of z is calculated. By this calculation, the state z itself is also changed to z^*. The process so described can be called a local change in the net. One can say that v_j is influenced by v_i and that v_i influences v_j, or one can say that v_j depends on v_i.

In a Bayes net, many local changes are possible. If a state, a special node i, and a special value for i are given, the i-th component can be changed. If this is done, the Bayesian machinery will begin to work. All the arrows beginning with node i are considered. Such an arrow leads to a node j. A local change can then be executed. In a bigger net, this leads to many local changes. If all these local changes are executed, the net is updated.

A Bayes net can be applied in many different real systems. In an application, the random variables $X_1,\ldots, X_n$ are the central part. In a given real system, a random variable X assigns to each event e (or to each sentence S) a value v: $X(e) = v$ (or $X(S) = v$). An event can be a real event or just a possible event that did not take place.

In a Bayes net, the events (or the sentences) "belong" to different nodes. A real system often contains many different events of different kinds and levels. Often this leads to a set or a list of events that can no longer be overlooked. At a sentence level, long lists of millions of sentences are difficult to understand. To analyze such "chaotic" lists of events or sentences, random variables are introduced. They should give some order to the set of events or sentences. In a Bayes net, two methods are used for this.

In the first method, events (or sentences) are collected into classes such that each class belongs exactly to one node of the net. It is said that events (or sentences) from such a class have the same dimension. Normally, such a dimension is expressed in natural language. An event or a sentence "has" or "belongs to" a special dimension. This is formulated by random variables. For each event or each sentence of a given dimension, a special random variable is used. Dimensions are distinguished using different random variables. A Bayes net has n nodes, where each node is responsible for the appertaining dimension. An event or a sentence can have a geometric dimension, a dimension of time, a dimension of a natural property, a social dimension, and so on.

In the second method, the events (or sentences), which have the same dimension, are further ordered by one distinct random variable X. To the events (or sentences) of this dimension, special values are assigned. In most cases, these values are numbers or binary symbols, for instance true and false or *t* and *f*. A random variable X will collect those events (or sentences) that are similar to one another. If e and e' are similar, they get the same value v: $X(e) = v = x(e')$.

The function of a Bayes net, therefore, can be described as follows. In a node, which is an input node of the net, the values of the appertaining random variable are known and given. From this node, all other nodes that lie in a path, beginning with the input node, will be updated. A flow of changes of values begins.

A Bayes net normally also has nodes in which a path cannot go further. These are called output nodes. A variable that is attached to a node that is not an input node can be rebound — i.e., the variable can be instantiated in different ways. All these possibilities can also be represented graphically.

This description of changes leads to probabilities and therefore to probability spaces. A Bayes net has one n-dimensional probability space and n one-dimensional probability spaces (Studentý 2010). A one-dimensional probability space belongs exactly to one node. All these spaces, normally, are just presupposed. The user should know these spaces and should know that an n-dimensional probability space can be constructed by n one-dimensional probability spaces.

3. Two examples of a Bayesian network

We would like to illustrate these rather abstract concepts with two examples — the first only superficial and informal, the second a little more detailed. One popular example, the "Asia model", is by (Lauritzen & Spiegelhalter 1988). In this model, a simple network is constructed to diagnose a doctor's new patients.

The Bayes net consists of nodes such as VisitAsia, Age, Smoking, or Tuberculosis. The direction of the links corresponds to causality. Each node represents a dimension that could be related to the patient's condition. For example, "Smoking" indicates that a patient is a smoker, and "VisitAsia" indicates whether the patient has recently been to Asia.

Probability relationships are indicated by the connections among the nodes. For example, an arrow goes from the Smoking node to both LungCancer and Bronchitis. Thus, for example, smoking increases both the likelihood of a patient developing bronchitis and the likelihood of them developing lung cancer, while age is only linked to the possibility of developing lung cancer.

Similarly, abnormalities on an X-ray of the lungs can be caused by either tuberculosis or lung cancer, while the likelihood of a patient suffering from breathlessness (dyspnea) is increased if the patient also has bronchitis or lung cancer.

Using this network, a doctor could determine how likely the patient is to have lung cancer or if the patient presents as a smoker and has a positive X-ray. In the second problem, a doctor might want to determine the most likely explanation that explains these symptoms — i.e., which conditions (e.g., tuberculosis, lung cancer, bronchitis) are most likely to have caused the symptoms.

In the second example, we look at a group of 10 people in two dimensions. The first dimension is health, and the second is poverty. We focus on a specific XY variant of the coronavirus for health and a generic description for poverty. A person is poor if he or she has relatively little money to live on and lives in a very small dwelling with several family members.

Expressed in normal language, no causal relationship between illness and poverty can initially be identified. Using a two-dimensional Bayes net, such a relationship can not only be recognized but also be described completely and precisely. The Bayes net contains, among other things, events from two dimensions, which are expressed by sentences. For this purpose, we form the list of names "Bob", "Mary", "Rose", "Peter",…, which we express generically by the "names" N_1,…, N_{10}; the variables N and N^* are to run over these persons.

The events we are interested in can be expressed in the first dimension by sentences such as "N_1 is sick", "N_2 is sick",…, "N_9 is healthy", "N_{10} is healthy". In the second dimension, the sentences are "N_1 is poor", "N_2 is poor",…, "N_9 is rich", "N_{10} is rich". We can see that events and sentences can be expressed in the same form.

In the Bayes net, these events and sentences are assigned to the two nodes. The two-dimensional Bayes net arises when the second node influences (depends on) the first. To formulate this, a two-dimensional probability space must be formed from the two sets of events and the associated one-dimensional prob-ability spaces. The sentences formulated above are linked to a hundred conjunctions of the type "N is sick and N^* is poor", "N is sick and N^* is rich", "N is healthy and N^* is poor", and "N is healthy and N^* is rich". Thus, we have formed a two-dimensional set E_2 of events (and sentences).

From these events, the random events can be constructed. In our example, we form the power set over E_2 and take all the elements from the power set and call them random events. This creates the natural form of the probability function p. A random event e has the probability α, $p(e) = \alpha$, if e has n elements and the number n is divided by the number m of events E (here $m = 100$).

The random events are further abbreviated in the example. The set of persons N for which "N is sick and for all N^*: N^* is poor or N^* is rich" (i.e., a random event), we denote by S; the set of persons for whom "N is healthy and for all N^*: N^* is poor or rich", we denote by H. Similarly, we introduce P for "N is poor and for all N^*: N^* is sick or healthy" and R for "N is rich and for all N^*: N^* is sick or healthy".

In the two-dimensional probability space, two random variables χ_1 and χ_2 are introduced. Values are as-signed to the events. These values can be not only numbers but also words or symbols. The random variable χ_1 has only two values v_1 and v_2. Instead of v_1, we write s (for sick), and instead of v_2, we write h (for healthy). Similarly, χ_2 has two values, w_1 and w_2; we denote these by p (poor) and r (rich).

The random variables are usually formulated in such a way that the random events that have the same value can be grouped into classes. All these classes are, by construction, random events. This allows the random events to be ordered optimally. Special random events are defined by the random variables. All events (sentences) that are assigned to the value v (i.e., $\chi_1(E) = v$) belong to the associated random event E. For example, all events (N_1 is sick and …), (N_2 is sick and …),…, (N_q is sick and …) are assigned to the value s (sick): χ_1(N_1 is sick and …) = χ_1(N_2 is sick and …),…, χ_1(N_q is sick and …) = s. The inverse image ${\chi^{-1}}_1(s)$ of the value s (sick) is, in other words, the set of events in which a person is sick. Thus, a random event {N_1 is sick and …, N_2 is sick and …, N_q is sick and …} is determined. This can be done in exactly the same way for all other values h (healthy), p (poor), and r (rich).

This allows us to form a Bayesian network with two nodes. A value v is assigned to certain random events e from E by χ_1. In the same way, a random event is assigned by χ_2 and by a value w. The value v determines the random event $E_1 = {\chi^{-1}}_1(v)$ and the inverse image of v under χ_1, and the value w determines the inverse image $E_2 = {\chi^{-1}}_2(w)$. This finally defines the probability for a state of this network. The state of the net is determined by the values v from the value set of χ_1 and by the values w from the value set of χ_2 and the net structure. In the example, this structure means that the second node is processed first and the first node second. Such a state is given by a special two-dimensional random event.

The probability is calculated as follows. In the first step, two values v and w are found that belong to the random event E. The corresponding inverse images ${\chi^{-1}}_1(v) = E_1$ and ${\chi^{-1}}_2(w) = E_2$ are determined.

In the second step, the average of the two inverse images is formed: $E_1 \cap E_2$. The random event $E_1 \cap E_2$ exists in the two-dimensional probability space by construction. This means that the number $p(E_1 \cap E_2)$ can be calculated; however, this can only be done in simple cases. In our example, the random variable X_1 is dependent on the random variable X_2 — that is, the following does not apply: $p(E_1 \cap E_2) = p(E_1) \cdot p(E_2)$.

The probability of a network state (i.e., a special random event) is defined in the example as follows: $p(E) = p(E_1 | E_2) \cdot p(E_2)$. Here, $p(E_1 | E_2) = p(E_1 \cap E_2) \cdot (1/p(E_2))$ is the conditional probability of E_1, conditioned by E_2.

For example, E_1 is the random event containing all events in which a person is sick, and E_2 is the random event containing all events in which a person is poor. With the numbers and symbols used here, the probability $p(E)$ of the network state at values s and p is calculated as follows, where $S = \{N$ is sick and ...$\} = X^{-1}{}_1(s)$, $P = \{N$ is poor and ...$\}$, and $E = S \cap P \cdot p(E) = p(S | P) \cdot p(P) = p(S \cap P) \cdot (1/p(P)) \cdot p(P)$.

If we use different values for the numbers of events, we see that the impact of poverty on sickness changes quantitatively. For example, if the number of sick people increases, the probability of the network state changes, and so does the influence of poverty on sickness. For example, if four people are sick, there are 40 two-dimensional events in which a sick person can be found, and if two people are rich, there are 20 events in which a rich person can be found. The probability in the network is then $p(e) = p(S | P) - p(P) = 40/100 = 2/5$. The set P of sick people is complementary to the set of rich people R: $P = 100 - 20 = 80$.

4. Events or sentences

A probability space can be formulated in two ways. In the first, the "classical" version, a probability space is based on events. In this account, events are distinguished into elementary events (or results) and random events, where a random event is a set of elementary events. To each random event, a probability (a number between 0 and 1) is assigned. The elementary events can be distinguished into those that really exist (here and now) and other elementary events that are only possible.

Probabilities are assigned to random events but not to elementary events. Only random events can have probabilities. Random events take place probabilistically; an elementary event cannot have a probability. An elementary event can only take place or not take place.

In the second formulation, a probability space is based on sentences (statements, propositions, terms, phrases). A probability function can be used in two ways. It can have one or two arguments. In the first variant, the probability function is written as p: $\mathbf{S} \rightarrow [0,1]$, $S \in \mathbf{S}$, $p(S) = \alpha$, and in the second version as $p^c(\_ | \_)$: $\mathbf{S} \times \mathbf{S} \rightarrow [0,1]$, $S, S^* \in \mathbf{S}$, $p^c(S | S^*) = \alpha$. In this variant, pairs $\langle S, S^* \rangle$ of sentences are assigned to a probability number. It is said that S has a probability α, conditional to S^*.

This conditional probability function can be defined in a classical probability space, but it can also be formulated independently of the classical approach. In a conditional probability space, all axioms can be formulated with the conditional probability function (Bernardo & Smith 1994).

In the event approach, the arguments of the random variable X are elementary events. The random variables can be seen as mediators between values and probabilities. A value v of X determines a set of elementary events, the inverse image $X^{-1}(v)$ of v, which has the probability $p(X^{-1}(v))$.

In the sentence approach, an argument will be transformed into a pair of sentences $\langle S, S^* \rangle$ so that to $\langle S, S^* \rangle$, two random variables X and X^* are assigned: $X(S) = v$ and $X^*(S^*) = v^*$. This notation is, however, not used. In the literature, especially in statistics, the sentences S and S^* are specialized to "equations" of the forms $X = v$ and $X^* = v^*$. Instead of $X(S) = v$, it is written as $X = v$. It is said that a sentence S has a probability conditioned by S^*, where S gets the value $X = v$ and S^* gets the value $X^* =$

v^*. Instead of $X(S|S^*)$, it is written as $p^c(X = v|X^* = v^*)$. This transformation leads to formal details that we will not describe here. Informally, in one position (or in several positions) of a sentence S, a part of the sentence is replaced by a variable — say, by y. The result is a formula $S[y]$. $p^c(X = v|S^*)$ says that the value v expresses how probable it is that a sentence of the form $X = v$ is true.

Sentences are believed by persons. Sentences have truth values, and beliefs have probabilities. With the help of beliefs, one can relate truth to probability (Hailperin 1984). How probable it is that a sentence is true (or false) is often asked. In this way, a sentence can not only be true or false. There are other possibilities: a sentence can be probable; it is true with a certain probability.

Beliefs are normally expressed by sentences or phrases from natural languages. In the sentence approach, a sentence can describe a belief of a person. In the event approach, one can say that an event represents a belief of a person. In central applications of Bayes nets, dependencies of beliefs are studied. A belief can be true or only probable to some degree; the probability of a belief is expressed in degrees.

Sentences and beliefs lead to the complex world of meaning. A sentence can interpret things, relations, sets of things, and sets of relations — and therefore events. This leads to a jungle of dependencies and influences.

In human thinking, "things" or "objects" are ordered by sets or collections. Different sets can comprise another set. This leads to many different levels—from basic things to metaphysical entities. In these complex collections of events, several special sets of events are distinguished. It is said that the events of such a set belong to one dimension. For instance, we can distinguish events that are connected with the term "intelligence" and other events that are connected with the legal term "fair". These events belong to different dimensions. The notion of dimension was clarified by Menger (1943).

The two versions of probability spaces belong to two world views, which are discussed without end. In our article, however, we will not participate in this discussion. We think that ontologically speaking, sentences are special kinds of events. This implies that the representation of a sentence and of a belief, which exists in the human brain, is an event.

In the following section, we formulate the models in the classical version.

5. The basic notions

n is an integer. n designs the number of dimensions. These dimensions can be understood in many ways. In applications, the notion of dimension is used not only for events (or sentences) but also for descriptions of events.

$\Omega_1,\ldots, \Omega_n$ are n sets of elementary events. Often the term result is used instead of the term "elementary event". We want to avoid the word "result" because this implies experiments from the natural sciences. $\Omega_1,\ldots, \Omega_n$ are base sets of (potential) models.

$A_1,\ldots, A_n$ are n sets of random events. A random event itself is a set of elementary events. For example, the random event "Rose likes her friends" is described by a list of elementary events: "Rose likes Peter", "Rose likes Bob", and so on.

$p_1,\ldots, p_n$ are n ("classical") probability functions. A probability function p_i has one argument E_i. To each random event $E_i \in A_i$, a probability α is assigned: $p_i(E_i) = \alpha$, p_i: $A_i \rightarrow [0,1]$.

From the sets $\Omega_1,\ldots, \Omega_n$, the Cartesian product $\Pi_{i\leq n}\Omega_i$ is constructed. We often abbreviate this product to $\mathbf{\Omega}$.

$X_1,\ldots, X_n$ are n random variables. Each random variable X_i is a function of the form X_i: $\Pi_{i\leq n}\Omega_i \rightarrow V_i$. Each set V_i, we call the set of values for X_i. We say that to each elementary event $e \in \Pi_{i\leq n}\Omega_i$, a value $v \in V_i$ is assigned: $X_i(e) = v$. Random variables are normally not one–one. A function value v can come from more than one argument: $\mathcal{X}_i(e) = v = \mathcal{X}_i(e')$, where $e \neq e'$. We comprise $\mathcal{X}_1,\ldots, \mathcal{X}_n$ into a list $\langle\mathcal{X}_1,\ldots, \mathcal{X}_n\rangle$, writing

$$\mathcal{X} = \langle\mathcal{X}_1,\ldots, \mathcal{X}_n\rangle.$$

The values are normally numbers (integers or real numbers) or terms such as 1 and 0 or true and false. In the second case, the random variable $\mathcal{X}_i$ is called binary. In infinite cases, the frequentist approach (von Mises 1981) to probabilities reaches its limits. There, often densities have to be used, and this leads to the full, formal machinery of probability theory and statistics.

The sets of values of these random variables form the basis for many models in which learning plays a role. A value represents a type or a trait of a property or an item of a specific area.

Using different random variables, the dimensions are distinguished. Even one event (or one sentence) can bear different dimensions. For example, a statement (and an event) can refer to color and noise at the same time: "This is a brown and noisy animal". Without using random values, distinctions must be expressed by the meaning of phrases that are parts of a statement. This leads to a lack of clarity. In a Bayes net, this problem is avoided because a random variable refers exactly to one appertaining dimension.

Formally, a set V_i is a part of the function $\mathcal{X}_i$. Therefore, it would not be necessary to insert this term V_i into the list of basic notions. However, in the description and application of Bayes nets, the sets $V_1,\ldots, V_n$ play a central role. For the structural completeness of the reconstruction, we must include these sets of values as base sets in the (potential) models. Also, we comprise the sets of values to a list $V = (V_i)_{i\leq n}$.

b is the influence relation. This relation lies in the core of Bayes nets. Formally, the nodes can be identified with random variables, as already discussed. The lines (or arrows) of the net, therefore, can be represented by pairs of random variables. We designate the set of these lines by b:

$$b \subset \{\mathcal{X}_1,\ldots, \mathcal{X}_n\} \times \{\mathcal{X}_1,\ldots, \mathcal{X}_n\}.$$

Formally, we should write $\langle\mathcal{X}_i, \mathcal{X}_j\rangle \in b$, but for better reading, we will transform the term $\langle\mathcal{X}_i, \mathcal{X}_j\rangle \in b$ into $\mathcal{X}_i b \mathcal{X}_j$. $\mathcal{X}_i b \mathcal{X}_j$ means that $\mathcal{X}_j$ influences $\mathcal{X}_i$. It is important to note that in the term "$\mathcal{X}_j$ influences $\mathcal{X}_i$", the indices are written from left to right, while in the term "$\mathcal{X}_i b \mathcal{X}_j$", the "stream" comes from the right to the left.

Apart from these basic notions, three further notions can be explicitly defined.

D is the distribution function.[1]

To formulate this function, we construct the Cartesian product $\Pi_{i\leq n}V_i = V_1 \times \ldots \times V_n$ of the sets of values of the random variables and typify D as a function as follows: D: $V_1 \times \ldots \times V_n \rightarrow [0,1]$. To each list of arguments $\langle v_1,\ldots, v_n\rangle$ a real number α is assigned, $D(v_1,\ldots, v_n) = \alpha$. $\langle v_1,\ldots, v_n\rangle$ is called a state of the Bayes net. This number expresses the probability of a state of the Bayes net.

In the following, we avoid double indices for symbols G by adding a double index i_j (or several indices) in angular brackets, for instance G[i_j] or G[i, r_i, $\xi(r_i)$]. For consistency, in some cases, we also write, for instance, G[i, ξ] if the indices are not double indices.

The described list of sets forms a potential model for Bayes theory, including a distribution function.

> x is an n-dimensional potential model of a Bayes net ($x \in BNp^n$) iff there exist $\Omega_1,\ldots, \Omega_n$, $V_1,\ldots, V_n$, $A_1,\ldots, A_n$, $p_1,\ldots, p_n$, $\mathcal{X}_1,\ldots, \mathcal{X}_n$, b, and D so that the following is true:

[1] In the statistics literature, the distribution function is often designated by the same symbol as for the probability function p. This will lead to misunderstandings if the reader is not a statistician. In probability theory, the symbol $p\mathcal{X}$ is used to distinguish a probability function p from the appertaining distribution $p\mathcal{X}$. In our article, we attempt to avoid all these possible misunderstandings and use different symbols.

1) $x = \langle \Pi_{i\leq n}\Omega_i, (V_i)_{i\leq n}, \mathbb{N}, [0,1], (A_i)_{i\leq n}, (p_i)_{i\leq n}, (X_i)_{i\leq n}, b, D\rangle$
2) $\mathbb{N}$ is the set of natural numbers, $n \in \mathbb{N}$ and $2 \leq n$
3) For all $i \leq n$, the following is true:
3.1) $\langle \Omega_i, A_i, p_i\rangle$ is a probability space
3.2) V_i is a non-empty, finite set
3.3) $X_i: \Pi_{i\leq n}\Omega_i \rightarrow V_i$
4) $\langle \Pi_{i\leq n}\Omega_i, \otimes_{i\leq n}A_i, \otimes_{i\leq n}p_i\rangle$ is an *n*-dimensional probability space[2]
5) $b \subset \{X_1,\ldots, X_n\} \times \{X_1,\ldots, X_n\}$
6) $D: V_1 \times \ldots \times V_n \rightarrow [0,1]$

We introduce the following abbreviations: $\Omega = \Pi_{i\leq n}\Omega_i$, $A = \otimes_{i\leq n}A_i$, $p = \otimes_{i\leq n}p_i$, $X = \langle X_1,\ldots, X_n\rangle$, and $V = \langle V_1,\ldots, V_n\rangle$. A potential model then has the following form:

$$x = \langle \Omega, V, \mathbb{N}, [0,1], A, p, X, b, D\rangle.$$

6. The models

The formulation of models of Bayes nets needs some auxiliary definitions. Though the notion of an n-dimensional probability space plays an important role in the Bayes nets, it is normally "outsourced". In Bayes nets, probability theory is treated as an auxiliary theory. We therefore put the *n*-dimensional probability space into the appendix.

For a random variable X_i, we define the inverse image of a value v:

(i) $X^{-1}{}_1(v) = \{e \in \Omega / X_i(e) = v\}$.

The inverse image of a value v is the set of all elementary events that have the same value v. These inverse images form random events. If the set of random events is not identical to the power set of elementary events, it must be assumed that all inverse images are also random events. With Definition (i), a random event can be written in different forms. We can write, for instance, the following:

$$p(X^{-1}{}_1(v_1) \cap \ldots \cap X^{-1}{}_n(v_n)) = p(\{e_1 \in \Omega / X_1(e_1) = v_1\} \cap \ldots \cap \{e_n \in \Omega / X_n(e_n) = v_n\}).$$

The distribution function D is defined as follows:

(ii) For all $v_1 \in V_1$ and … and for all $v_n \in V_n$:

$$D(v_1,\ldots, v_n) = p(X^{-1}{}_1(v_1) \cap \ldots \cap X^{-1}{}_n(v_n)).$$

In this definition, the inverse images $X^{-1}{}_1(v_1),\ldots, X^{-1}{}_n(v_n)$ are identical with random events $E_1,\ldots, E_n$, and these can be represented by sets $\{e_{11},\ldots, e_{1}s_{1}\},\ldots, \{e_{n1},\ldots, e_{n}s_{n}\}$ of elementary events. At the sentence level, the cardinalities of sets $\{e_{i1},\ldots, e_{i}s_{i}\}$ can be determined, for instance by counting. In these cases, the cardinalities can be used as absolute frequencies. In finite models in which all these sets are random events, these frequencies lead directly to probabilities, namely, to the probabilities of the random events $E_1,\ldots, E_n$, $E_i = X^{-1}{}_i(v_j)$, $p(E_i) = p(X^{-1}{}_i(v_j))$. On the right side of Definition (ii), these random events

[2] The constructions of Cartesian products $\Pi_{i\leq n}\Omega_i$ and $\Pi_{i\leq n}V_i$, of σ-algebras $\otimes_{i\leq n}A_i$, and of *n*-dimensional probability functions (or probability measures) $\otimes_{i\leq n}p_i$ are not formulated here. These standard definitions are found in the literature (e.g., Loéve 2017). It is interesting to note that in the standard literature about Bayes nets, these notions are even not mentioned.

are "concatenated" by the formal operation of intersection $E_1 \cap \ldots \cap E_n$, and $D(v_1,\ldots, v_n)$ expresses the probability of this intersection.[3]

With these definitions, the independence of random variables $\mathcal{X}_i$ and $\mathcal{X}_j$ in a n-dimensional probability space W is formulated.

(iii) $\mathcal{X}_i$ is independent of $\mathcal{X}_j$ iff
for all $v_1 \in V_1$, $v_r \in V_i$, $v_s \in V_j$ and for all $v_n \in V_n$, it holds that[4]

$$p(\mathcal{X}^{-1}{}_1(v_1),\ldots, \mathcal{X}^{-1}{}_i(v_r) \cap \mathcal{X}^{-1}{}_j(v_s),\ldots., \mathcal{X}^{-1}{}_n(v_n)) =$$
$$p(\mathcal{X}^{-1}{}_1(v_1),\ldots, \mathcal{X}^{-1}{}_i(v_r),\ldots., \mathcal{X}^{-1}{}_n(v_n)) \cdot p(\mathcal{X}^{-1}{}_1(v_1),\ldots, \mathcal{X}^{-1}{}_j(v_s),\ldots., \mathcal{X}^{-1}{}_n(v_n)).$$

Informally, we can say that $\mathcal{X}_i$ depends on $\mathcal{X}_j$ if the probability coming from a value v_j cannot be determined independently of the probability, coming from v_i. This means that one event can influence another event, even if both events "live" in different dimensions. This definition can be extended to conditional independence, which plays an important role today (Endres & Augustin 2016).

In simple cases, Definition (iii) can be expressed by a set of equations. For all values $v \in V_i$ for $\mathcal{X}_i$ and for all $v' \in V_j$ for $\mathcal{X}_j$ and the appertaining random events $H = \mathcal{X}^{-1}{}_i(v)$ and $E = \mathcal{X}^{-1}{}_j(v')$: $p(H \cap E) = p(H) \cdot p(E)$. The product of two beliefs about two random events H and E is identical with the belief of the intersection of the random event $H \cap E$, and this is true for all combinations of random events.

In such a simple case, Bayes theory leads to a completely symmetric equation:

$$(p(H \cap E)/p(E)) \cdot p(E) = p(H) \cdot (p(H \cap E)/p(H))$$

where H is a random event describing a hypothesis and E describes a datum.

In a probability space, this equation is normally expressed by a conditional probability function p^c, p^c: $A \times A \rightarrow [0,1]$. $p^c(H|E)$ expresses the probability of H conditional on E or, in other words, the probability of H — but only if E is true. In statistics, the term $p(H)$ is called the prior probability of H, $p(H|E)$ the posterior probability of H, conditional on E (or the posterior probability of H, assuming that E is true), $p(E|H)$ the likelihood of H given E, and $p(E|\neg H)$ the likelihood of $\neg H$ given E. Normally, this notation is further abbreviated. If the application is well known, it is just spoken about "the" posterior of H and "the" prior of H.

A further definition deals with the marginal distribution $D[r, \xi]$ of random variables $\mathcal{X}[i_1],\ldots, \mathcal{X}[i_r]$ in D, where $\{\mathcal{X}[i_1],\ldots, \mathcal{X}[i_r]\}$ is a subset of $\{\mathcal{X}_1,\ldots, \mathcal{X}_n\}$. $D[r, \xi]$ distributes only the random variables $\mathcal{X}[i_1],\ldots, \mathcal{X}[i_r]$, while the other random variables $\mathcal{X}_j$ ($\mathcal{X}_j \notin \{\mathcal{X}[i_1],\ldots, \mathcal{X}[i_r]\}$) are neutralized (see, for instance, Bernardo & Smith 1994).

(iv) $D[r, \xi](v[\xi(1)],\ldots, v[\xi(r)]) = \sum Q(\sigma,1) \ldots \sum Q(\sigma, n)\ D(w_1,\ldots, w_n)$, where for all $\sigma \in \{1,2\}$
and all $q \leq n$:
if $\sigma = 1$ and $\neg\exists j \leq n\ (q = \xi(j))$, then $Q(\sigma, q)$ refers to w_q, and it holds
that $w_q \in V_q$,
and
if $\sigma = 2$ and $\exists j \leq n\ (q = \xi(j))$, then $Q(\sigma, q)$ refers to w_q, and it holds
that $w_q = v_q$.

The set IN of input nodes and the set OUT of output nodes for $\{\mathcal{X}_1,\ldots, \mathcal{X}_n\}$ can be defined as follows:

(v) $\mathcal{X}_i \in IN$ iff $\neg\exists \mathcal{X}_j \in \{\mathcal{X}_1,\ldots, \mathcal{X}_n\}\ (\mathcal{X}_i \neq \mathcal{X}_j \wedge \mathcal{X}_i b \mathcal{X}_j)$ and

---

[3] At the sentence level, these events can be transformed into a conjunction. This conjunction represents the global state of the net at a given time, and $D(v_1,\ldots, v_n)$ expresses the probability of this state described by the conjunction. In the Bayes nets, such conjunctions are further analyzed.

[4] This formulation is sloppy, but we want to avoid x-ing out the indices, which leads to rather long definitions.

$X_i \in OUT$ iff $\neg\exists X_j \in \{X_1,\ldots, X_n\}$ $(X_i \neq X_j \wedge X_j b X_i)$.

For potential models $x \in BNp^n = \langle \Omega, V, \mathbb{N}, [0,1], A, p, X, b, D\rangle$, $i \leq n$ and $X_i \in X$, we define the following:

(vi) $EN(i)$ is an elementary net for X_i in x iff there exist r_i and ξ_i, so that

1) $EN(i) = \langle \Omega, V[\xi_i(1)],\ldots, V[\xi_i(r_i)], \mathbb{N}, [0,1], A, p, X[\xi_i(1)],\ldots, X[\xi_i(r_i)], b_i\rangle$
2) $1 \leq r_i \leq n$
3) ξ_i: $\{1,\ldots, r_i\} \rightarrow \{1,\ldots, n\}$ is one–one (injective)
4) $\xi_i(1) = i$ (and therefore $X[\xi_i(1)] = X_i$)
5) for all j with $2 \leq j \leq r_i$, it holds $X_i b X[\xi_i(j)]$
6) for all $j \leq r_i$ and all $X' \in X$: if $X' \neq X[\xi_i(j)] \neq X_i$ and $X' \neq X_i$, then $\neg(X_i b X')$ and $\neg(X' b X[\xi_i(j)])$
7) for all j with $2 \leq j \leq r_i$ and for all $s \leq n$, it holds that if $X[\xi_i(j)] \neq X_s$ and $X[\xi_i(j)] b X_s$, then $s \notin \{\xi_i(1),\ldots, \xi_i(r_i)\}$
8) $b_i = \{\langle X_i, X[\xi_i(2)]\rangle,\ldots, \langle X_i, X[\xi_i(r_i)]\rangle\}$
9) for all u, $v \leq r_i$ with $1 < u$ and $1 < v$: $X[\xi_i(u)]$ is independent of $X[\xi_i(v)]$.

In the appendix, we show that for each X_i, an elementary net exists. Finally, we define the local distribution function $D[r_i, \xi_i]$ for X_i and $EN(i)$ in x.

(vii) $D[r_i, \xi_i]$ is the local distribution function for X_i and $EN(i)$ in x:

$D[r_i, \xi_i]$: $V[\xi_i(1)] \times V[\xi_i(2)] \times \ldots \times V[\xi_i(r_i)] \rightarrow [0,1]$, where
$D[r_i, \xi_i]$ is the marginal distribution of $X[\xi_i(1)],\ldots, X[\xi_i(r_i)]$ in D.

This function distributes only a given random variable X_i and the random variables that influence X_i directly.

For an elementary net $EN(i)$ for X_i in x, we call X_i the head of $EN(i)$ and $\langle X[\xi_i(2)],\ldots, X[\xi_i(r_i)]\rangle$ the list of parents of X_i. This list can be empty.

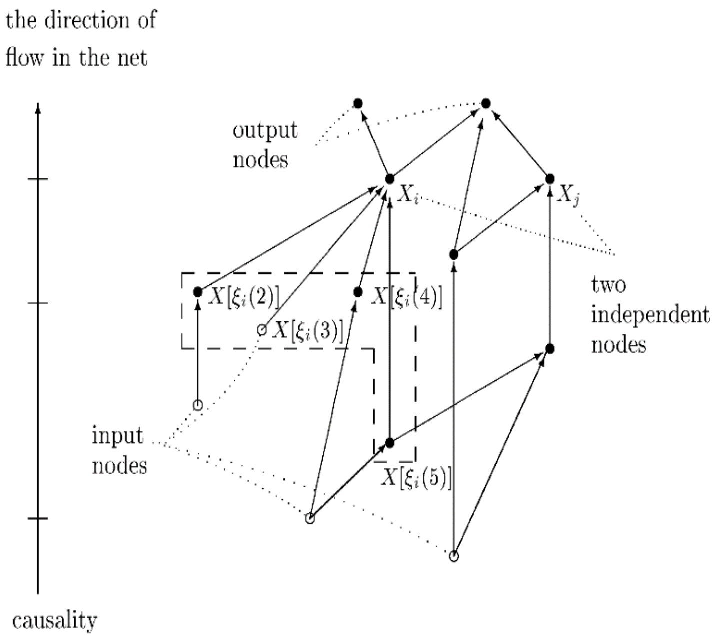

Figure 1: A model and an elementary net for X_i.

Here is an example: $5 < n = 13$, $I = 1$, $r = r_i = 5$, and $\xi = \xi_i$ is the identity function.

$D[5, \xi](v[\xi(1)],\dots, v[\xi(5)]) = \sum Q(\sigma,1) \dots \sum Q(\sigma,5)\ D(v_1,\dots, v_5, v_6,\dots, v_{13})$, where for all $u \leq 5$, $Q(\sigma, u)$ refers to v_u. If $\sigma = 2$ and $u \leq 5$, then $v_u = v[\xi(u)]$, and if $\sigma = 1$ and $5 \leq u \leq 13$, then $v_u \in V[\xi(u)]$.

With these definitions, a Bayes net can be characterized by the following hypotheses (1)–(6) (see the definitions for Ω, A, and p above):

x is an n-dimensional model for Bayes nets ($x \in BN_n$) iff there exist $\Omega_1,\dots, \Omega_n$, $V_1,\dots, V_n$, $A_1,\dots, A_n$, $p_1,\dots, p_n$, $X_1,\dots, X_n$, b, and D so that

1) $x = \langle \Omega, V_1,\dots, V_n, \mathbb{N}, [0,1], A, p, X_1,\dots, X_n, b, D\rangle$ and $x \in BNp^n$
2) there exist $Y, Y^* \in \{X_1,\dots, X_n\}$ so that $Y \neq Y^*$ and YbY^*
3) for all $Y \in \{X_1,\dots, X_n\}$: not (YbY)
4) for all $Y_1, Y_2, Y_3 \in \{X_1,\dots, X_n\}$, it holds that
   if Y_1bY_2 and Y_2bY_3, then Y_1bY_3
5) for all $k \leq n$ and for all $Y[i_1],\dots, Y[i_k] \in \{X_1,\dots, X_n\}$:
   if $Y[i_1]bY[i_2]$ and $Y[i_2]bY[i_3]$ and … and $Y[i_k - 1]bY[i_k]$, then
   this is not true: $Y[i_k]bY[i_1]$
6) for all $v_1 \in V_1$ … for all $v_n \in V_n$:
   $D(v_1,\dots, v_n) = \Pi_{1\leq i\leq n}\ D[r_i,\xi_i](v_i, v[\xi_i(2)],\dots, v[\xi_i(r_i)])$.

Hypothesis (1) expresses the structure of the model and describes the probability space. Hypothesis (2) says that there exist at least two different random variables (nodes) so that Y^* influences Y. Hypothesis (3) expresses that no random variable influences itself, (4) requires that the influence relation is transitive, and (5) says that in a Bayes net, the influence relation cannot lead to circles.

This should be understood as follows. In a Bayes net, paths can be defined. A path is a sequence $Y_1,\dots, Y_s$ of nodes so that $Y[1]bY[2]$ and $Y[2]bY[3]$ … and $Y[k - 1]bY[k]$. A path is circular if the last node $Y[k]$ also influences the first one: $Y[1]bY[k]$. Based on the hypotheses, a unique direction of flow is defined. If Y influences Y^*, the hypotheses imply that Y^* cannot influence Y, and this is also true for nodes that are not directly related.

The central hypothesis (6) says that the distribution function D can be analyzed as a product of all local distribution functions $D[r_i,\xi_i]$, $i \leq n$. Exactly the random variables $X[\xi_i(j)] = X_j$ influence X_i.

A local distribution function can have two forms. In the first form, the head of X_i is independent of all other random variables. For the head of X_i, there are no parents for X_i. In the second form, there are random variables (namely, parents) that influence X_i.

In the literature about Bayes nets, other notations are used, such as

$$p^cX(v_1,\dots, v_k \mid v_{k+1},\dots,v_n) = \Pi\ p^cX(v \mid par(v[j_1],\dots, v[j_s])), \text{ where}$$

v is a value of X_i and $par(v[j_1],\dots, v[j_s])$ designates a list of values of the parent nodes for X_i. The exact relation between indices i and j_q ($= \xi_i(q)$) are not seen anymore. More detailed formulations could, of course, be added, such as

$$\Pi\ p^cX(v \mid par(v[j_1],\dots, v[j_s])) = \Pi_{1\leq i\leq n}p^cX(v_i \mid par(v[j_1],\dots, v[j_s])) = \Pi_{1\leq i\leq n}p^cX(v[\xi_i(1)] \mid par(v[\xi_i(2)],\dots, v[\xi_i(r_i)])).$$

A Bayes net always has some input nodes; this follows from the hypotheses. The input nodes cannot be influenced. In these nodes, new data will be inscribed, and as there is a clear direction, this data can (and will) flow through the net. In the literature, the direction of a Bayes net is often depicted without marking a global direction, but normally, the input nodes are depicted at the front of the net. We always depict Bayes nets so that the input nodes lie at the bottom and all the other nodes that are influenced are found upward.

A node X_i that is not an input or an output node lies in the inner, theoretical part of the net. From such a node, we can look upward and determine those nodes that are influenced by X_i, and we can look downward and determine other nodes that influence X_i directly.

We mention that the order of the random variables in a net, which is given by the indices 1,..., n, has no additional meaning. If the indices are varied, the resulting net is isomorphic to the original one.

Bayes nets can be used for formal means and for empirical applications. In the latter, we can order the random variables so that the dependent ones are written first and the independent ones last.

7. Intended systems

From the subjectivist view, probability is a personal affair. Probabilities are entities found in the cognitive systems of persons. In an intended system, there are two ways to determine whether an event has a probability. In the first way, the actor is asked a question by a researcher, and the actor responds and expresses a belief of an event under discussion.[5]

In such cases, if many answers are given, frequencies could be determined. In the second way, however, the actor itself does not or cannot respond. For instance, a robot or a small child, at first, must learn before it or he can respond. In these cases, frequentist methods are difficult to use.

Methods of determination of probabilities (and beliefs) and influences of beliefs are found in many intended systems in different domains, even in legal and moral situations, in medicine, and in formal applications. Even a mathematician has beliefs about certain assumptions, which he uses in a deduction. All these cases can be regarded as intended systems for Bayes theory.

An intended system can often be seen as a process by which an actor learns something effectively. A person should learn, for example, to cross a street at a place where a traffic light exists. Several sentences can be used for the appertaining events: “The person distinguishes colors and forms”, “Her legs function”, “This machine is a traffic light”, “Cars drive fast or slowly or stand still”, etc. Not all sentences in a description are atomic. In these cases, they can be further analyzed so that at the end, a description consists only of atomic sentences (and perhaps negations). In the example, the conjunction of colors and forms can be resolved into two sentences (“The person distinguishes colors” and “The person distinguishes forms”), and the adjunction “fast or slowly or stand still” can be resolved into three sentences. Colors and velocities can be distinguished further in different ways and in different degrees. Colors can be expressed, for instance, by “green”, “red”, etc. and velocities by adjectives such as “very fast”, “fast”, “slow”, “very slow”, and “stand still”. In this way, sentences can be transformed to conjunctions of atomic statements so that each elementary sentence belongs to a different dimension. In a given situation, the traffic light can, for instance, be only “red” or “green” so that in a description of the dimension for color only, one of the two sentences can be found. Similarly, a moving car can have “just now” only one of several velocities.

An intended system y for Bayes theory can be normed so that it can be described with the notions introduced above. The intended system then has the form of a partial model:

$$y = \langle \Omega^*, V^*_1,\ldots, V^*_{n^*}, \mathbb{N}, [0,1], A^*, p^*, X^*_1,\ldots, X^*_{n^*}, b^*, D^* \rangle.$$

Looking at the components of a potential model $x = \langle \Omega, V_1,\ldots, V_n, \mathbb{N}, [0,1], A, p, X_1,\ldots, X_n, b, D \rangle$, we determine that some of the components can be partially determined. Some facts from the intended system can be collected. The parts Ω^*, A^*, p^*, $V^*_1,\ldots, V^*_{n^*}$, $X^*_1,\ldots, X^*_{n^*}$ of the components of Ω, A, $V_1,\ldots, V_n$, $X_1,\ldots, X_n$ of a potential model can be listed.

[5] The researcher and the actor can, of course, be identical.

In an application, some sentences for elementary events $e \in \Omega^*$ and some values $v \in V^*_i$ of the appertaining random variable X^*_i are known. Such values are often expressed in a qualitative way, not by real numbers. For instance, one can "see" that one list of sentences has a different form from a second list of sentences. In this way, sentence classes are formed and normalized to sets of atomic sentences — i.e., inverse images of values of X_i. If the sets of atomic sentences are not too small, it is possible to assign values to such classes. In this way, even random variables X_i can be partially determined. In some applications, some frequencies of sentences of the same form can also be determined so that first probability values $p^*(E)$ can be included in the intended system. All this can be expressed verbally.

Looking at these sentences and at the appertaining real system, various dimensions can be distinguished. This is done by dividing the set of sentences into different classes that belong to the appertaining dimensions. In this way, the number n of dimensions should be determined. If n was fixed, it is not necessary to use additional dimensions. Therefore, the number n^* of dimensions will be identical to the theoretical number n.

The last components b and D are more difficult to relate to facts. They refer to items of a more abstract level. b relates random variables to other random variables, but how can such relations be determined, and how can values of D be empirically determined? These two components lead to problems of the structure of nets and to the problem of theoreticity (Sneed 1971).

In many theories, it is possible to divide the basic notions of the theory into theoretical and non-theoretical terms. We cannot go into the details here (Balzer 1985). We assume that parts Ω^*, $V^*_1, \ldots, V^*_{n^*}$, A^*, p^*, $X^*_1, \ldots, X^*_{n^*}$ of y can be determined by facts, while b and D are difficult to link directly with facts if the hypotheses of Bayes theory are not used. The non-theoretical components, which contain facts, are listed, while parts of b and D cannot be determined without using the hypotheses.

At this point, partial models y are completed. We can ask whether a partial model y can be embedded (see the introduction) into a model x. We say that a component of x is theoretical iff the appertaining set is empty. In our case, this means that b^* and D^* are theoretical.

For Bayes theory, we can formulate the empirical claim in a generalized version (Balzer, Lauth & Zoubek 1993), if we assume that an "open" set **I** of intended systems exists. In the core variant, for each partial model y, which is also an intended system, there exists a model x so that y is embedded into x:

For all $y \in \mathbf{I}$ exists an $x \in \mathbf{M}$ so that y is embedded into x.

This claim should further be generalized so that an embedding holds only approximately.

We can ask, in other words, whether an influence relation b and a distribution function D exist so that y can be embedded into a model x. Normally, this will not be the case. Some data will not fit the model, but using a simple or complex statistical method, it is often possible to embed y into x approximately.

Instead of going further into the formal details, we will give three examples in which Bayes models are used. In the first example, a computer learns to land an airplane of a given type in a certain airport. A large amount of data must be observed and stored. In the second example, the computer learns to identify signatures. A large amount of data is collected from many different pictures of real signatures. The pictures are analyzed into pixels so that the pixels are distributed into a thousand (or more) small boxes. A given list of basic graphical forms must be available at the beginning. In each box, one finds one of the few basic graphical forms that can be determined by the computer. The computer will decide whether the content of a box that was observed and is received is identical to one of the basic "blueprint" forms.

In a third application of Bayes theory, a child learns to obey a given practical rule. The child and the mother observe many situations in which the child wants to cross a street down which many cars are driving. The mother tells the child to look first to the left and then to the right. If no car is coming from the left and if the same is true at the right, the child could go. Otherwise, the child should wait. After several learning situations, the child has learned this rule.

In these examples, by looking at real systems, different dimensions can be found. Different events are investigated, and a large amount of data is collected. In some way, their dimensions really exist. In the third example, we can, for instance, distinguish a dimension of geometrical procedures, a dimension of traffic, a dimension of danger, and a dimension of the effectiveness of learning. The number of these dimensions must be determined by the researcher. Such determinations are rather informal at the moment; they are created by the scientists, who look at the real systems and the sentences available.

8. Two final remarks

Different approaches to probability have been under discussion for many decades. In the objectivist view, the value of a probability function refers, at least in simple cases, to some real events or “things” that exist outside of a human body. In the subjectivist view, the value of a probability function refers to the belief of a human person. How beliefs from different persons become coherent is another story that is not told in Bayes theory — but Bayes models can occur in this story.

In our reconstruction, the models are based on events, not on sentences. Formally speaking, both versions are not far away from each other. Sentences can be constructed formally from events. In the other direction, we have no clear ideas on how to construct every event from sentences. Can all events that are communicated by humans also be expressed by sentences? For instance, we find mathematical entities in a Borel space in probability theory, which are related to other entities. However, there are no ways to construct all these entities. Also, some events can be imagined but cannot be communicated by a person.

The problem with probability, therefore, does not lie in the formulation of events or sentences. The problem lies in the world views, which are expressed by languages, and in the way of determining probabilities. In the classical version, in simple cases, a probability is determined by relative frequencies of elementary events. This is possible if several elementary events are at hand. Also, in a Bayes net, it is possible to investigate events that are similar to one another. Additionally, in this case, events and event classes can be determined, and frequencies can be calculated.

We think that the difference between using events and using sentences lies in the way of distinguishing the sets of intended systems. For applications in which only very few events are available in a system, frequencies are difficult to determine. Such a system is not included in the set of intended systems of the theory if the theory is used in the objectivist way. In the subjectivist way, however, real systems are also investigated, for which frequencies practically cannot be determined.

The second remark is that the theory described here forms the basis of a theory net (Balzer, Moulines & Sneed 1987, Chap. IV). Three specializations have already been investigated extensively. The first concerns methods of elimination of paths in a net. In a net, often different paths can be used to reach a given node. If such paths are not used for other streams, some of them can be eliminated. The influence relation b is further characterized so that independencies between nodes are minimized — that is, all the possible flows in the net will take the best possible path (Lauritzen 1982). The second specialization concerns time. In the basic variant of Bayes nets, time is not present. The description of a flow of information would refer to time points. Of course, time can be added to the models. This can be done by forming sequences of Bayes models with special properties. In a third branch of this theory net, only Markov processes are investigated.

9. Appendix

D1) W is a probability space $\langle\Omega, [0,1], A, p\rangle$ iff

1) Ω ist a set and $A \subseteq \wp(\Omega)$
2) p: $A \rightarrow [0,1]$
3) $\Omega \in A$ and Ω is not empty
4) If $E \in A$, then $E^c \in A$ (E^c designs the complement of the set E)
5) For each sequence $(E_i)_i = 1, 2, 3,\ldots$ with $E_i \in A$, it holds that $E_1 \cup E_2 \cup E_3 \cup \ldots \in A$
6) $p(\Omega) = 1$
7) For each sequence $E_1, E_2, E_3 \ldots \in A$ with pairwise disjoint members, it is true that $p(\cup i\ E_i) = \sum i = 1, 2, 3,\ldots\ p(E_i)$

Let $n > 2$, and for all $i \leq n$, let $W_i = \langle\Omega_i, [0,1], A_i, p_i\rangle$ be a probability space.

$W = \langle\Omega, [0,1], A, p\rangle$ is the n-dimensional probability space of $W_1,\ldots, W_n$, where $\Omega = \Pi_{i\leq n}\Omega$, $A = \otimes_{i\leq n}A_i$, and $p = \otimes_{i\leq n}p_i$. The random events $E \in \otimes_{i\leq n}A_i$ are sets of n-tuples of elementary events $e_i \in \Omega_i$ $(i \leq n)$. These random events E are constructed basically from n-dimensional rectangles, and from these with unions, intersections, and complements, a minimal set of random events is defined such that the axioms for the probability space are true for this set.

Theorem:
Let $x = \langle\Omega, (V_i)_{i\leq n}, \mathbb{N}, [0,1], A, \mathrm{p}, (\mathcal{X}_i)_{i\leq n}, b, D\rangle \in BNp^n$ and $i \leq n$ be given. We show that there exist r_i and ξ_i so that

1) $1 \leq r_i \leq n$,
2) ξ_i: $\{1,\ldots, r_i\} \rightarrow \{1,\ldots, n\}$ is injective, $\xi_i(1) = i$, and
3) $(\{\mathcal{X}[\xi_i(1)],\ldots, \mathcal{X}[\xi_i(r_i)]\} \times \{\mathcal{X}[\xi_i(1)],\ldots, \mathcal{X}[\xi_i(r_i)]\}) \subset b$.

Proof:
We define a function φ_i: $\{0, 1, 2,\ldots, n\} \rightarrow \cup_{0\leq m<n} (\{\mathcal{X}[1],\ldots, \mathcal{X}[n]\}_{m+1})$ inductively as follows:
$m = 0$: $\varphi_i(0) = \langle\mathcal{X}[i]\rangle$
$m = 1$: $\varphi_i(1) = \langle\mathcal{X}[i], \mathcal{X}[i,j_1]\rangle$ if there exists j_1 so that $j_1 = \mu_s F_0(s, \mathcal{X}[1],\ldots, \mathcal{X}[n])$
or
$= \langle\mathcal{X}[i], 0\rangle$, in all other cases.

Here, $\mu_s F_0(s, \mathcal{X}[1],\ldots, \mathcal{X}[n])$ means that s is the smallest number from $\{1,\ldots, n\}$ so that $F_0(s, \mathcal{X}[1],\ldots, \mathcal{X}[n])$ is true, and $F_0(s, \mathcal{X}[1],\ldots, \mathcal{X}[n])$ is an abbreviation for '$1 \leq s \leq n$; also, there exists $\mathcal{X}[i]b\mathcal{X}[s]$, and there exists no $\mathcal{X}[q] \in \{\mathcal{X}[1],\ldots, \mathcal{X}[n]\}$ so that ($\mathcal{X}[q]b\mathcal{X}[s]$ and $\mathcal{X}[i]b\mathcal{X}[q]$ and $q \neq i$)'.

The step of induction, from m to $m+1$, proceeds as follows:
Let $1 \leq m < n$, and we suppose that

$\varphi_i(m) = \langle\mathcal{X}[i], \mathcal{X}[i, j_1],\ldots, \mathcal{X}[i, j_m]\rangle$ was defined.
$\varphi_i(m+1) = \langle\mathcal{X}[i], \mathcal{X}[i,j_1],\ldots, \mathcal{X}[i,j_m], \mathcal{X}[i,j_{m+1}]\rangle$ if there exists j_{m+1} so that $j_{m+1} = \mu_s F_1(s, \mathcal{X}_1,\ldots, \mathcal{X}_n)$
or
$= \langle\mathcal{X}[i], \mathcal{X}[i, j_1],\ldots, \mathcal{X}[i, j_m], 0\rangle$ in all other cases.

$\mu_s F_1(s, \mathcal{X}[1],\ldots, \mathcal{X}[n])$ means that s is the smallest number from $\{1,\ldots, n\}$ so that $F_1(s, \mathcal{X}[1],\ldots, \mathcal{X}[n])$ is true, and $F_1(s, \mathcal{X}[1],\ldots, \mathcal{X}[n])$ is an abbreviation for '$1 \leq s \leq n$; also, there exists $\mathcal{X}[i]b\mathcal{X}[s]$, and there exists no $\mathcal{X}[q] \in \{\mathcal{X}[1],\ldots, \mathcal{X}[n]\} \setminus \{\mathcal{X}[i], \mathcal{X}[i, j_1],\ldots, \mathcal{X}[i, j_m]\}$ so that ($\mathcal{X}[q]b\mathcal{X}[s]$ and $\mathcal{X}[i]b\mathcal{X}[q]$ and $q \neq i$)'.

Therefore, the ordering for $\varphi_i(n) = \langle X[i], X[i, j_1],\ldots, X[i, j_{n-1}]\rangle$ is determined. We eliminate all components '0' from the list $\varphi_i(n)$ and obtain the reduced list $\langle X[i], X[i,2],\ldots, X[i,r_i]\rangle$, where $X[i] = X[\xi(1)]$. Therefore, r_i and ξ_i exist.

References

Abreu, C., Lorenzano, P. and C. U. Moulines (2013), "Bibliography of Structuralism (1995–2012 and Additions)", *Metatheoria* 3: 1-36. https://doi.org/10.48160/18532330me3.91

Balzer, W. (1985), "On a New Definition of Theoreticity", *Dialectica* 39: 127-145. https://doi.org/10.1111/j.1746-8361.1985.tb01251.x

Balzer, W., Lauth, B. and G. Zoubek (1993), "A Model for Science Kinematics", *Studia Logica* 52: 519-548. https://doi.org/10.1007/bf01053258

Balzer, W., C. U. Moulines and J. D. Sneed (1987), *An Architectonic for Science*, Dordrecht: D. Reidel. https://doi.org/10.1007/978-94-009-3765-9

Balzer, W. and J. D. Sneed (1977/78), "Generalized Net Structures of Empirical Theories I + II", *Studia Logica* 36: 195-211, and 37: 167-194. https://doi.org/10.1007/BF02121266

Bernardo, J. M. and A. F. M. Smith (1994), *Bayesian Theory*, New York: Wiley. https://doi.org/10.1002/9780470316870

Bremaud, P. (1999), *Markov Chains*, Berlin: Springer Verlag. https://doi.org/10.1007/978-1-4757-3124-8

Cox, R. (1947), "Probability, Frequency, and Reasonable Expectation", *American Journal of Physics* 14(1): 1-13. https://doi.org/10.1119/1.1990764

Endres, E. and T. Augustin (2016), "Statistical matching of discrete data by Bayesian networks", in Antonucci, A., Corani, G. and C. P. de Campos (eds.), *Proceedings of the Eighth International Conference on Probabilistic Graphical Models, Vol. 52 of Proceedings of Machine Learning Research*, Lugano: PMLR, pp. 159-170.

Hailperin, T. (1984), "Probability Logic", *Notre Dame Journal of Formal Logic* 25: 198-212. https://doi.org/10.1305/ndjfl/1093870625

Kolmogorov, A. N. (1950), *Foundations of the Theory of Probability*, New York: Chelsea Publishing.

Lauritzen, S. L. (1982), *Lectures on Contingency Tables*, Denmark: University of Aalborg Press, 2nd ed.

Lauritzen, S. L. and D. J. Spiegelhalter (1988), "Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems", *Journal of the Royal Statistical Society* 50(2): 157-224. https://doi.org/10.1111/j.2517-6161.1988.tb01721.x

Loéve, M. (2017), *Probability Theory*, New York: Dover Publications, Mineola, revised 3rd ed. https://doi.org/10.1007/978-1-4684-9464-8

Menger, K. (1943), "What Is Dimension?", *The American Mathematical Monthly* 50(1): 2-7. https://doi.org/10.1080/00029890.1943.11991313

Michie, D. (ed.) (1979), *Expert Systems in the Micro-Electric Age*, Edinburg: Edinburgh University Press.

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Francisco, California: Morgan Kaufman Publishers Inc., revised 2nd ed. https://doi.org/10.1016/C2009-0-27609-4

Renyi, A. (1970), *Foundations of Probability Theory*, San Francisco: Holden-Day. (Translation of Renyi, A. 1962. *Wahrscheinlichkeitsrechnung*, Berlin: VEB Deutscher Verlag der Wissenschaften.)

Schiffer, A. (1987), *Remnants of Meaning*, Massachusetts: Cambridge. https://doi.org/10.2307/2185491

Sneed, J. D. (1971), *The Logical Structure of Mathematical Physics*, Dordrecht: Reidel. https://doi.org/10.1007/978-94-010-3066-3

Sokolowski, J. A. and C. M. Banks (eds.)(2010), *Modeling and Simulation Fundamentals: Theoretical Underpinning and Practical Domains*, Hoboken, NJ: John Wiley & Sons.

Studentý, M. (2010), *On Probabilistic Conditional Independence Structures*, London: Springer.

von Mises, R. (1981), *Probability, Statistics and Truth*, New York: Dover Publications, 2nd ed.