Basic Properties of Class Hierarchies regarding Probability Distributions

Wenyun Quan Patrick van Bommel Peter J. F. Lucas

Institute for Computing and Information Sciences
Radboud University Nijmegen, The Netherlands
{w.quan,pvb,peterl}@cs.ru.nl

Abstract

Concepts from object orientation have been applied to many fields to facilitate solving complex real-world problems. Medicine is an example of such a complex field, where, however, also the modeling of uncertainty is of major importance. It is our belief that object orientation can also play a role in the medical field to make representing and reasoning with uncertain knowledge easier. However, there is little known about how ideas from object orientation affect the specification and use of probability distributions. In this paper it is studied in what way structured probabilistic models can be organized in class hierarchies. We will provide a theoretical foundation of probabilistic models with object orientation, which are called probabilistic class hierarchies. This is expected to offer a basis for the modeling of complex problems, such as those in medicine, from which the examples used in this paper come.

Keywords: class hierarchies, probabilistic networks, object-oriented models, disease modeling.

1 Introduction

The aim of this paper is to explore ideas around class hierarchies as known from object-oriented languages in probabilistic graphical models, and in particular we are interested in whether this makes specification of probabilistic graphical models easier. Most other object-oriented probabilistic languages have so far refrained from using ideas of class hierarchies, with the exception of [2], [4] and [6]. However, in these papers using class hierarchies for the specification of probability distributions has no implications for the associated probability distributions. As in these previous papers, we will use Bayesian networks to represent probability distributions.

In our paper, we extend standard class hierarchies toward *probabilistic class hierarchies* by defining probability distributions for classes in hierarchies. It is first explored in what way class hierarchies can be extended from a probabilistic point of view. It will appear that the knowledge represented in a class hierarchy poses insufficient constraints to obtain a probabilistic class hierarchy straight away. However, by associating probability distributions to classes it will become clear that it is possible to obtain sensible constraints to these probability distributions. The wish to obtain such results was the main motivation why we started with this research. Some basic properties of probabilistic class hierarchies are examined, which lead us to a promising method to deal with complex problems, such as in medicine, the example domain of this paper.

The paper is organized as follows. The research discussed in this paper was motivated by the difficulty in the medical field to represent uncertain knowledge of multiple diseases, as is illustrated in Section 2. In Section 3, some basic notions from standard class hierarchies and probability theory are recalled. The new concept of a probabilistic class hierarchy, which is a standard class hierarchy combined with probability distributions, is introduced in Section 4. This section also contains the important properties we have been able to derive. Further conclusions and future work are discussed in Section 5.

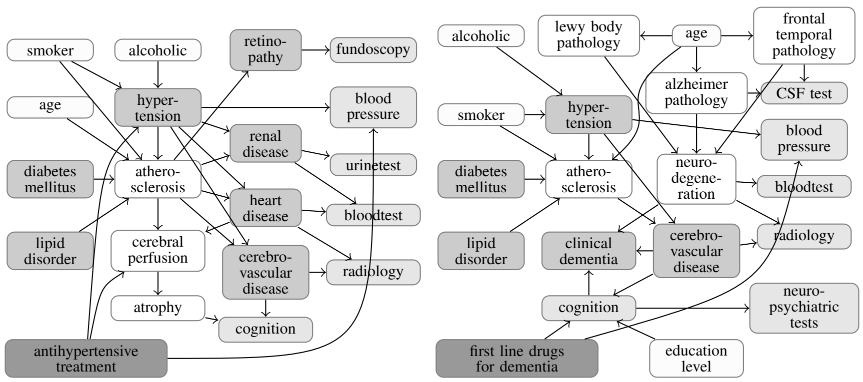

Figure 1: The graphical interpretations of the clinic guidelines for two diseases: hypertension (left) and dementia (right).

2 Motivating Example

During the past decade, the presence of multiple disease, also known as *multimorbidity*, has become an increasingly important health-care problem in the modern, western society. In particularly, multimorbidity is the norm in the elderly patient. Unfortunately, medical knowledge whether contained in clinical guidelines or medical textbooks, is mostly organized around single disorders. That is why there are no guarantees whether an elderly patient receives appropriate treatment. Probabilistic graphical models support modeling knowledge of the interaction between multiple diseases, both in terms of manifestations and treatment actions and their effects, and also support the modeling of the temporal aspects of clinical guidelines. For instance, Figure 1 shows the graphical representations of knowledge about hypertension and dementia, with the vertices representing variables and arcs representing interactions between variables. We use different gray levels to categorize those variables. However, which interactions exist between these two disorders is not clear even though the models share some variables. Furthermore, a disadvantage of using a graphical representation of multiple diseases that does show such interactions is that it will become very complicated. The same holds for the associated probability distributions. There is, thus, a strong need to add levels of abstractions and means for reuse to Bayesian network representation in medicine, hence the notion of probabilistic class hierarchies developed in this paper.

3 Preliminaries

In this section, we review the standard theory of class hierarchies [1] and probability distributions [3].

3.1 Class Hierarchies

Classes and the relationships between classes are the starting point for any form of object orientation. Let U be a set of all classes, called the *universe*. We write $c' \preceq c$ to indicate that class c' is a *subclass* of class c. Note that $\preceq$ possesses the properties of a *partial order*, i.e. reflexivity, transitivity and antisymmetry. A *proper subclass relation* $\prec$ is a *weak partial order* that, in contract to $\preceq$, possesses irreflexivity and asymmetry.

Classes are used to define generic objects. These objects have properties that are shared by their instances. Such properties are defined by means of *attributes*, basically functions from classes to other classes. For class c with $n \in \mathbb{N}$ attributes, we use the notation $c(a_1 : c_1, \ldots, a_n : c_n)$ meaning that a_i is an attribute of class c having c_i as a type, or a class as we will say. We call them *class structures*. Classes may have no attribute at all, and then we write $c()$. A_c is used to denote the associated set of the attributes of c, e.g. $A_c = \{a, a'\}$ if we have $c(a : c', a' : c'')$. For a class with no attribute, i.e. $c()$, A_c is defined as $A_c = \{\epsilon\}$, where ϵ denotes the *empty attribute*. An attribute may occur in multiple classes, but is unique within any class.

A *class hierarchy* H is defined as a triple $H = (C, \preceq, S)$, where S is the set of the class structures associated with the classes in C, and the partial order $\preceq$ represents subclass relations. Given two classes $c, c' \in C$ in a class hierarchy H with $c' \preceq c$, then $c''' \preceq c''$ holds for all attribute $a \in A_c$ with the class structures $c(a : c'')$ and $c'(a : c''')$.

In the end, classes and class structures are meant to act as templates for instantiation, yielding *instances*. Instances of a class are meant to represent *individual* objects in the real world. For instances of objects, we use the notation $i(a = i')$ given class structure $c(a : c')$, where i is an instance of class c and i' is an instance of c'. The instance-class relationship is denoted by $i \ll c$.

A *path* is a sequence σ of attributes, which yields meaningful class information. The concatenation of the empty attribute ϵ with any attribute yields again the attribute, i.e. $a.\epsilon = a$. The *paths of attribute* a with respect to class c, denoted by $\rho_c\,(a)$, are defined as:

$$\rho_c(a) = \begin{cases} \{a.\sigma \mid \sigma \in \rho_{c'}(a')\} & \text{if } c(a : c') \in S; \\ \{\epsilon\} & \text{otherwise.} \end{cases}$$

For instance, if we extend the above-mentioned class structures with $c''(a' : c'''')$, then the paths of a with respect to c is $\rho_c(a) = \{a, a.a'\}$. We can categorize the paths of an attribute with respect to a class into *extendable paths* and *non-extendable paths*. All paths with infinite length are extendable paths, the paths with finite length can be extendable, but can also be non-extendable. If the class structure of c'' is $c''(a' : c)$ instead of $c''(a' : c'''')$, then the set of the paths of a with respect to c is infinite, i.e. $\rho_c(a) = \{a, a.a', a.a'.a, a.a'.a.a', \ldots\}$. We call the attributes having classes with no attribute *basic attributes*. The non-extendable paths are the paths terminating with basic attributes. The *final classes* of an attribute a with respect to c are the classes of the last attributes on the finite paths of a. Obviously, all final classes of non-extendable paths have no attributes.

The *interpretation* of a class *without* attributes, denoted by $\Gamma(c)$, translates to a set, e.g. $\Gamma(\text{nat}) = \mathbb{N}$, i.e., the set of natural numbers. The *domain of an attribute* a *with respect to class* c, denoted by $\delta_c(a)$, is the Cartesian product of the interpretations of the non-extendable paths' final classes of this attribute with respect to c. The interpretation of a class *with* attributes is equivalent to the Cartesian product of the domains of all attributes with respect to this class, formally, $\Gamma(c) = \bigtimes_{a \in A_c} \delta_c(a)$. Given the classes c and c' with $c' \preceq c$ and A_c as the set of the attributes of c, then $\delta_{c'}(a) \subseteq \delta_c(a)$ holds for all $a \in A_c$, as a basic property of class hierarchies. Given two classes c and c' in the same class hierarchy, then $c' \preceq c \Rightarrow \Gamma(c') \subseteq \Gamma(c)$ always holds. Therefore, $c' \prec c \Rightarrow \Gamma(c') \subset \Gamma(c)$ also holds. In this case, there exists at least one attribute, e.g., a with $\delta_{c'}(a) \subset \delta_c(a)$ We call such attributes the *characteristic attributes* of c regarding c'. We say that classes c and c' are *mutually exclusive*, if $\Gamma(c) \cap \Gamma(c') = \varnothing$ holds. The union of two mutually exclusive classes c and c' is denoted by $c \,\dot\cup\, c'$. Given a class c and a finite set of its subclasses C, the subclass set C is a *partition* of class c, if the equation $c = \dot\bigcup_{c' \in C} c'$ holds.

3.2 Probability Theory and Bayesian Networks

The set of all possible outcomes of an experiment is called the *sample space*, denoted by Ω. An *event* E is a subset of the sample space Ω, but not all the subsets of Ω are events. The collection of events can be considered as a subcollection $\mathcal{F}$ of the set of all subsets of Ω. We call a collection of subsets of Ω a σ*-field*, denoted by $\mathcal{F}$, if it includes the empty set, the union of all subset of $\mathcal{F}$ and the complement of all member of $\mathcal{F}$. A *real-valued random variable* is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$. We use upper case letters, e.g., X, to represent a random variable, and lower case letters, e.g., x, to represent a specific value of X. The random variable X is called *discrete* if it takes values in some countable subset of $\mathbb{R}$. The discrete random variable X has *probability mass function* $f : \mathbb{R} \to [0, 1]$ given by $f(x) = P(X = x)$. A *cumulative distribution function* of a random variable X is a function $F : \mathbb{R} \to [0, 1]$ given by $F(x) = P(X \leq x)$. The random variable X is called *continuous* if its distribution function can be expressed as $F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u$ with $x \in \mathbb{R}$ for some integrable function $f : \mathbb{R} \to [0, 1]$, called the *probability density function* of X. Let E, E' and E'' be three events, such that $E = E' \,\dot\cup\, E''$. Now, from $E' \subseteq E$, it follows that $P(E') \leq P(E)$, as $P(E) = P(E') + P(E'')$. This basic property offers a natural start, but not more than that, to give a probabilistic interpretation to identical attributes of two subclasses of a given class (see below).

Given two sets of the random variables X and Y, defined on the same probability space, the *joint distribution* of X and Y defines the probability of events in terms of both X and Y: $P(X, Y)$. Given the joint distribution of X and Y, the *marginal distribution* of X is calculated by summing (for discrete random variables) or integrating (for continuous random variables) the joint probability distribution over Y. If $P(Y) > 0$, then the *conditional probability* that X may occur given that Y has occurred is defined as $P(X \mid Y) = P(X, Y)/P(Y)$. Sets of the random variables X and Y are called *conditionally independent* given the set of random variables Z if the conditional probability of X given Y and Z is the same as the conditional probability of X given Z, i.e. $P(X \mid Y, Z) = P(X \mid Z)$. A conditional probability can be inversed using *Bayes' rule*: $P(X \mid Y) = P(Y \mid X)P(X)/P(Y)$.

A *Bayesian network* [5] is a probabilistic graphical model, denoted by $\mathcal{B} = (G, X, P)$, where $G = (V, E)$ is a directed acyclic graph (DAG) with a set of vertices V and an edge relation $E \subseteq V \times V$, X denotes a set of variables, and P is the joint probability distribution of X, factored according to the structure of G. There is a 1–1 correspondence between sets V and X. The joint probability distribution of a Bayesian network can be calculated from the conditional probabilities using the chain rule.

4 Probabilistic Class Hierarchies

In the previous sections, we reviewed class hierarchies and probability theory. For standard class hierarchies, the property $\Gamma(c') \subseteq \Gamma(c)$ if $c' \preceq c$ is seen as the essential. In order to cope with the representation of uncertainty in such an object-oriented framework, we need to somehow associate classes and attributes with probability distributions. This will lead us towards *probabilistic class hierarchies*. In the following section, we will introduce ideas and examples on how class hierarchies can be associated to probability distributions. Subsequently, formal properties are studied.

4.1 Initial Ideas on Probabilistic Class Hierarchies

Koller and Pfeffer [4] use a so-called Bayesian network fragment in their object-oriented Bayesian networks, OOBNs for short, to describe the probabilistic relations between the attributes of a class. In the OOBN approach, a class with probability distributions is represented graphically by a DAG $G_c = (V_c, E_c)$, where V_c represents the set of vertices labeled with the random variable corresponding to attributes of class c, i.e. A_c. $E_c \subseteq V_c \times V_c$ is the set of directed edges between the vertices, mirroring direct probabilistic interactions between attributes. The graph is a compact representation of the set of *independencies* that hold for the associated probability distribution P_c of the random variables corresponding to the attributes of c, that is why we call the graph an *independency structure*. In this paper, we use an attribute a of a class c as index to the corresponding random variables, i.e. X_a, which are denoted by the attributes' names initialed with capital letters in independency structures instead; and x_a denotes a possible value of X_a.

The random variable corresponding to an attribute is a real-valued function mapping the domain of this attribute with respect to a class into the real line $\mathbb{R}$, i.e., $X_a : \delta_c(a) \to \mathbb{R}$. Thus, due to the ordering of the real numbers, a total order is defined on the domain of each attribute. Let X_a and $X_{a'}$ be the random variables corresponding to the attributes a and a', respectively. Attribute a' is a *parent* of a with respect to class c, if there exists an arc from the vertex labeled with $X_{a'}$ to that labeled with X_a. We denote the set of parents of attribute a with respect to class c by $\pi_c(a)$. Let x_a and $x_{a'}$ be a value of the random variables X_a and $X_{a'}$, respectively, with $x_a, x_{a'} \in \mathbb{R}$. If X_a is a discrete variable, the conditional probability distribution of a with respect to class c is equal to the probability mass function of X_a, i.e.

$$P_c(X_a = x_a \mid \{X_{a'} = x_{a'} \mid a' \in \pi_c(a)\}) = f_{c.a}(x_a \mid \{x_{a' \in \pi_c(a)}\}).$$

If X_a is a continuous variable, the conditional probability distribution of a with respect to class c is equal to the cumulative distribution function of X_a, i.e.

$$P_c(X_a \leq x_a \mid \{X_{a'} = x_{a'} \mid a' \in \pi_c(a)\}) = F_{c.a}(x_a \mid \{x_{a' \in \pi_c(a)}\}).$$

The joint probability distribution of class c is equal to the product of the conditional distributions of all attributes of c. Similar to ordinary Bayesian networks, the independency structure defines the factorization of the joint distribution $P_c(\{X_a \mid a \in A_c\})$ associated with the graph concerning class c.

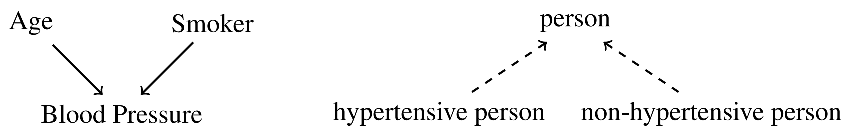

Figure 2: The independency structure of class person (left) and the structure of probabilistic class hierarchy person (right).

Example 4.1. *First, we give an example of a joint probability distribution associated to a class. The class structure of class* person *is defined as* $\text{person}(\text{age} : 0..127, \text{smoker} : \text{boolean}, \text{blood pressure} : 70..200)$, *where class* boolean *is without attributes and is thus interpreted as* $\Gamma(\text{boolean}) = \{\text{true}, \text{false}\}$. *The independency structure G_p is shown in Figure 2(left); p is taken as an abbreviation of 'person'; similar abbreviations are used for 'age', 'smoker', and 'blood pressure'. Let X_a and X_s be the random variables corresponding to the attributes* age *and* smoker*. The conditional probability distribution of the continuous random variable X_b corresponding to attribute* blood pressure *with respect to class* person *is equal to the conditional cumulative distribution function of X_b with respect to class* person*, i.e., $P_p(X_b \leq x_b \mid X_a = x_a, X_s = x_s) = F_{p.b}(x_b \mid x_a, x_s)$. Therefore, the joint probability distribution of the random variables corresponding to the attributes of class* person *is equal to the joint cumulative distribution function of the random variables of class* person*, i.e., $P_p(X_a = x_a, X_s = x_s, X_b \leq x_b) = P_p(X_b \leq x_b \mid X_a = x_a, X_s = x_s)P_p(X_a = x_s)P_p(X_s = x_s) = F_p(x_a, x_s, x_b)$.*

For understanding the remainder of this paper, it is also important to realize that the joint probability distribution for a class c, i.e., $P_c(\{X_{a \in A_c}\})$ should be interpreted as a probability distribution *conditioned* on a random variable of classes Y, i.e., $P_c(\{X_{a \in A_c}\}) = P(\{X_{a \in A_c}\} \mid Y = c)$, where $Y = c$ means that the context of the distribution is class c.

Let us assume that a class c can be partitioned into two mutually exclusive subclasses c' and c''. What we would like to examine is under which conditions the subclass relation carries over to a subclass relationship with probability distributions, where $c' \preceq c$ may have some implications with respect to the relationship between P_c and $P_{c'}$. The fact that c' and c'' are both subclasses of c may also have consequences. We continue with the example.

Example 4.2. *Let us partition class* person *into two mutually exclusive probabilistic subclasses 'hypertensive person' and 'non-hypertensive person', with the class structures:*

> $\text{hypertensive person}(\text{age} : 0..127, \text{smoker} : \text{boolean}, \text{blood pressure} : 95..200)$, *and*
> $\text{non-hypertensive person}(\text{age} : 0..127, \text{smoker} : \text{boolean}, \text{blood pressure} : 70..95)$.

We assume that these two subclasses have both the same attribute set as class person*. In this case, the attribute* blood pressure *is the only characteristic attribute of class* person *regarding class* hypertensive person *and class* non-hypertensive person*. The structure of the class hierarchy is shown in Figure 2(right). We will use similar abbreviations as before for these additional classes, here '*hp*' and '*nhp*'.*

Since the attributes age*,* smoker *and* blood pressure *are translated to random variables X_a, X_s and X_b, the original domains are transformed to real numbers, allowing us to compare the associated (joint) probability distributions. Assuming for the time being that the independence information is the same for the three classes, the joint probability distributions are $P_c(X_a, X_s, X_b) = P_c(X_b \mid X_a, X_s)P_c(X_a)P_c(X_s)$ for $c \in \{\text{p}, \text{hp}, \text{nhp}\}$. The joint cumulative distribution function of the random variables of c is $F_c(x_a, x_s, x_b) = \Sigma_{v \leq x_s} \int_{-\infty}^{x_a} \int_{-\infty}^{x_b} f_c(u, v, w)\mathrm{d}u\mathrm{d}w$, where $f_c(x_a, x_s, x_b) = f_{c.b}(x_b \mid x_a, x_s)f_{c.a}(x_a)f_{c.s}(x_s)$ is the joint probability density function of the random variables of c. $f_{c.b}(x_b \mid x_a, x_s)$ is the conditional density distribution function of X_b, $f_{c.a}(x_a)$ and $f_{c.s}(x_s)$ are the marginal distributions of X_a and X_s, respectively.*

As specifications of joint mass and density functions, possibly as factors in a Bayesian network, are sufficient to obtain any probability distribution concerning any Boolean expression, they seem to offer a good foundation for the comparison of joint probability distributions of classes, illustrated in the following example. Recall that joint mass and density functions are always defined on the real numbers, and this renders it possible to compare such functions to each other.

Example 4.3. *We continue elaborating the idea. We assume that attribute blood pressure obeys normal distributions in all classes, denoted by* $X_b \sim \mathcal{N}(\mu_{c.b}, \sigma^2_{c.b})$ *with* $c \in \{\mathrm{p}, \mathrm{hp}, \mathrm{nhp}\}$*. The blood pressure of a* hypertensive person *is likely higher than the blood pressure of a* person*, and it is certainly higher than that a* non-hypertensive person*. Now let us assume that the variances of the normal distributions with respect to those classes are identical, i.e.,* $\sigma_{p.b} = \sigma_{hp.b} = \sigma_{nhp.b}$*. As the class set* $\{\mathrm{hp}, \mathrm{nhp}\}$ *is a partition of class* person*, then the following inequality of their means must hold:* $\mu_{nhp.b} \leq \mu_{p.b} \leq \mu_{hp.b}$*. Therefore, the relation of the conditional cumulative distribution functions for* blood pressure *with respect to these three classes is* $F_{hp.b}(x_b \mid x_a, x_s) \leq F_{p.b}(x_b \mid x_a, x_s) \leq F_{nhp.b}(x_b \mid x_a, x_s)$ *for all* $x_a, x_s, x_b \in \mathbb{R}$*. The attributes* age *and* smoker *are the parents of the characteristic attribute* blood pressure *(see Figure 2(left)) and their probability distributions may also vary in the subclasses in comparison with their distributions in class* person*. The following inequalities of the probability distributions of the attributes* age *and* smoker *with respect to the different classes are assumed to hold, i.e.* $F_{hp.a}(x_a) \leq F_{p.a}(x_a) \leq F_{nhp.a}(x_a)$ *and* $F_{hp.s}(x_s) \leq F_{p.s}(x_s) \leq F_{nhp.s}(x_s)$*, for all* $x_a, x_s \in \mathbb{R}$*.*

In this example, it was possible to order the probability distributions of the random variables corresponding to the attributes, although these orders did not all respect the subclass relations between the associated classes.

Example 4.4. *As a conclusion of the previous examples,* $F_{hp}(x_a, x_s, x_b) \leq F_p(x_a, x_s, x_b) \leq F_{nhp}(x_a, x_s, x_b)$ *holds for all* $x_a, x_s, x_b \in \mathbb{R}$*, with* $\mathrm{hp}, \mathrm{nhp} \preceq \mathrm{p}$*, yielding an order that does not respect the subclass relation between* nhp *and* p *but suggests the existence of derivable constraints.*

4.2 Probability Distributions Constrained by Class Hierarchies

In the previous section, we used some examples to illustrate the idea that knowledge expressed by a class hierarchy may have implications for associated probability distributions. Now, we give a formal definition of probabilistic class hierarchies which will allow us to prove properties of class hierarchies that always hold. Note that we assume that in probabilistic class hierarchies, the attribute sets of a class and its subclasses are identical.

Definition 4.5. *A* probabilistic class hierarchy *is defined as a tuple* $H_P = (C, \preceq, S, X, P)$*, where* X *denotes the set of random variables corresponding to the attributes defined in* S*,* P *is the set of the joint probability distributions of each class* $c \in C$*, and the definitions of* C*,* $\preceq$ *and* S *remain the same as for standard class hierarchies* H*.*

Example 4.6. *The probabilistic class hierarchy defined in the examples of Section 4.1 contains the set of random variables, i.e.* $X = \{X_a, X_s, X_b\}$*, and its associated set of the joint probability distributions with respect to all different classes, i.e.* $P = \{P_p(X_a, X_s, X_b), P_{hp}(X_a, X_s, X_b), P_{nhp}(X_a, X_s, X_b)\}$*.*

We gave an example of a probabilistic class hierarchy in Section 4.1 using the assumption that the conditional probability distributions of the random variable corresponding blood pressure with respect to all three classes obey the normal distributions with the same variance. In fact, the following lemma shows that an important, general property holds for any probability distribution.

Lemma 4.7. *Given the classes* c*,* c' *and* c'' *with* $c', c'' \preceq c$ *and* $c = c' \mathbin{\dot\cup} c''$*, let* X_a *be the random variable corresponding to attribute* $a \in A_c$*, then from* $P_{c'}(X_a) \leq P_c(X_a)$ *it follows that* $P_{c''}(X_a) \geq P_c(X_a)$*, and vice versa, for* X_a *either discrete or continuous.*

Proof. We first define the probability distribution of classes, denoted by $P(Y)$. Let $c', c'' \preceq c$ be subclasses of c that partition c. Clearly, $P(Y = c \mid E) = P(Y = c' \mid E) + P(Y = c'' \mid E)$, for any event E (including the empty one). We use the definition of probability distributions conditioned on classes: $P_c(X_a) = P(X_a \mid Y = c)$. We wish to prove that from $P_{c'}(X_a) \leq P_c(X_a)$ it follows that $P_{c''}(X_a) \geq P_c(X_a)$, and vice versa (where also $\geq$ and $\leq$ can be interchanged).

Using Bayes' rule: $P(X_a \mid Y = c) = P(Y = c \mid X_a)P(X_a)/P(Y = c) \propto P(Y = c \mid X_a)/P(Y = c)$. Similarly, $P(X_a \mid Y = c') \propto P(Y = c' \mid X_a)/P(Y = c')$, and $P(X_a \mid Y = c'') \propto P(Y = c'' \mid X_a)/P(Y = c'')$. Important is to realize that in the proportionality expressions we left out the same constant: $P(X_a)$. Now, let us assume that $P_{c'}(X_a) \leq P_c(X_a)$, then, as probabilities are positive numbers

$$
\begin{aligned}
P_{c'}(X_a) \leq P_c(X_a) \quad &\Leftrightarrow \quad P(Y = c' \mid X_a)/P(Y = c') \leq P(Y = c \mid X_a)/P(Y = c) \\
&\Leftrightarrow \quad \alpha = P(Y = c' \mid X_a)/P(Y = c \mid X_a) \leq \beta = P(Y = c')/P(Y = c).
\end{aligned}
$$

Note that $0 \leq \alpha \leq \beta \leq 1$. Thus, $P(Y = c' \mid X_a) = \alpha P(Y = c \mid X_a)$, from which follows that $P(Y = c'' \mid X_a) = P(Y = c \mid X_a) - P(Y = c' \mid X_a) = P(Y = c \mid X_a) - \alpha P(Y = c \mid X_a) = (1 - \alpha)P(Y = c \mid X_a)$; and $P(Y = c') = \beta P(Y = c)$, from which it follows that $P(Y = c'') = P(Y = c) - P(Y = c') = P(Y = c) - \beta P(Y = c) = (1 - \beta)P(Y = c)$.

We now need to compare $P_c(X_a)$ and $P_{c''}(X_a)$. We have:

$$
P(X_a \mid Y = c'') \propto \frac{P(Y = c'' \mid X_a)}{P(Y = c'')} = \frac{(1 - \alpha)P(Y = c \mid X_a)}{(1 - \beta)P(Y = c)}.
$$

From $0 \leq \alpha \leq \beta \leq 1$ it follows that $(1 - \alpha)/(1 - \beta) \geq 1$, and $P(X_a \mid Y = c'') \geq P(X_a \mid Y = c)$. Hence, $P_{c''}(X_a) \geq P_c(X_a)$. The proof for $P_{c'}(X_a) \geq P_c(X_a)$ is similar. □

Lemma 4.7 also holds if the partition of a class has an arbitrary size. Because of the equivalence between the probability distribution of a random variable and its distribution function, we obtain the following, *weaker* corollary, that, however allows to express a generic property.

Corollary 4.8. *Given a class c which is partitioned into a set of classes C, if X_a is the random variable corresponding to attribute a, the following inequality of conditional cumulative distribution functions holds for all $a \in A_c$ and $x_a, x_{a'} \in \mathbb{R}$:*

$$
\min_{c' \in C}\{F_{c'.a}(x_a \mid \{x_{a' \in \pi_{c'}(a)}\})\} \leq F_{c.a}(x_a \mid \{x_{a' \in \pi_c(a)}\}) \leq \max_{c' \in C}\{F_{c'.a}(x_a \mid \{x_{a' \in \pi_{c'}(a)}\})\}.
$$

Furthermore, this property also applies to the joint cumulative distribution functions, so that

$$
\min_{c' \in C}\{F_{c'}(\{x_{a \in A_c}\})\} \leq F_c(\{x_{a \in A_c}\}) \leq \max_{c' \in C}\{F_{c'}(\{x_{a \in A_c}\})\}
$$

holds for all $a \in A_c$ and $x_a \in \mathbb{R}$.

A related question is whether we can actually derive probabilistic information from the subclasses for a given class. The following lemma concerns expected values of random variables for a partition of a given class c.

Lemma 4.9. *Let C be a partition of class c and $\alpha_{c'} = P(Y = c')$ be the weight of the subclass $c' \in C$ regarding class c, with $\sum_{c' \in C} \alpha_{c'} = 1$. X_a denotes the random variable corresponding to attribute $a \in A_c$ and the random variable Y determines the context of the distribution. The properties of the* expected values *of X_a with respect to class c, denoted by $\mathbf{E}_c[X_a]$, are:*

- $\mathbf{E}_c[X_a] = \sum_{c' \in C} \alpha_{c'} \mathbf{E}_{c'}[X_a]$, *if X_a is discrete and $\mathbf{E}_{c'}[X_a] = \sum_{x_a} x_a P_{c'}(X_a = x_a)$ is the expected value of X_a with respect to class c', where $P_{c'}(X_a = x_a)$ is the probability distribution of X_a with respect to c';*
- $\mathbf{E}_c[X_a] = \int_{-\infty}^{\infty} x_a f_c(x_a)\, \mathsf{d}x_a = \sum_{c' \in C} \alpha_{c'} \int_{-\infty}^{\infty} x_a f_{c'}(x_a)\, \mathsf{d}x_a$, *if X_a is continuous, $f_c(x_a)$ and $f_{c'}(x_a)$ are the density functions of X_a with respect to class c and class c', respectively.*

Proof. Since C is a partition of c, $P(Y = c) = \sum_{c' \in C} P(Y = c') = 1$. Therefore, $\sum_{c' \in C} \alpha_{c'} = 1$. If X_a is a discrete random variable, then

$$
\begin{aligned}
\mathbf{E}_c[X_a] &= \sum_{x_a} x_a P(X_a = x_a \mid Y = c) = \sum_{x_a} x_a P(X_a = x_a \mid Y = c) P(Y = c) \ \langle \text{as } P(Y = c) = 1 \rangle \\
&= \sum_{x_a} x_a P(X_a = x_a, Y = c) = \sum_{c' \in C} \sum_{x_a} x_a P(X_a = x_a, Y = c') \ \langle \text{as } c = \dot{\bigcup}_{c' \in C} c' \rangle \\
&= \sum_{c' \in C} \sum_{x_a} x_a P(X_a = x_a \mid Y = c') P(Y = c') = \sum_{c' \in C} \mathbf{E}_{c'}[X_a] P(Y = c') \\
&= \sum_{c' \in C} \alpha_{c'} \mathbf{E}_{c'}[X_a] \ \langle \text{as } \alpha_{c'} = P(Y = c') \rangle.
\end{aligned}
$$

The proof for a continuous random variable X_a is analogous. $\square$

Example 4.10. *We modify Example 4.3 by introducing the class* ordinary person *and the class* hypotensive person *to replace class* non-hypertensive person*. In this case, the class* person *is partitioned into three subclasses, i.e., class* hypertensive person*, class* ordinary person *and class* hypotensive person*, in a probabilistic class hierarchy, renamed to* hype*,* ordp *and* hypo*, respectively. Let X_a be the random variable corresponding to attribute* age*, we assume $\mathbf{E}_{hype}[X_a] = 75$, $\mathbf{E}_{ordp}[X_a] = 50$ and $\mathbf{E}_{hypo}[X_a] = 40$. Let α_c with $c \in \{\text{hype}, \text{ordp}, \text{hypo}\}$ be the weight of c corresponding to its superclass* person*, we assume $\alpha_{hype} = 0.2$, $\alpha_{ordp} = 0.5$ and $\alpha_{hypo} = 0.3$. Therefore, the expected value of X_a with respect to class* person *is*

$$\begin{aligned}\mathbf{E}_p[X_a] &= \alpha_{hype} \cdot \mathbf{E}_{hype}[X_a] + \alpha_{ordp} \cdot \mathbf{E}_{ordp}[X_a] + \alpha_{hypo} \cdot \mathbf{E}_{hypo}[X_a] \\ &= 0.2 \times 75 + 0.5 \times 50 + 0.3 \times 40 = 52.\end{aligned}$$

5 Conclusion and Future Work

The aim of this research was to determine whether the concept of class hierarchy could be extended in a probabilistic fashion. There was clearly no straightforward way to extend class hierarchies in this way, but what has become clear is that definitions of class hierarchies do can have implications for probability distributions in the sense of the provision of constraints. We examined in this paper some basic properties of probabilistic class hierarchies by defining probability distributions in terms of classes and their attributes. We believe these properties may lead us to complete the existing approaches to probabilistic object-oriented modeling and establish a robust foundation for such models, which are considered to provide a promising method to solve a complex problem, especially in the medical field.

There are still questions in this research that remain unanswered. We will focus on two of such questions for future research. First, although we discovered the basic relations between the probability distributions of a class and the members of its partition, we wish to move further by looking for more specific relationships between actual distributions. Second, we only defined probabilistic class hierarchies with identical attribute sets for each class and its subclasses. While correct from a formal point of view, this property may make it hard to use such a probabilistic object-oriented model in practice. However, what we do expect is that such models may facilitate the analysis of databases, as we here start with some additional knowledge about relationship between instances found in the data. Our ultimate aim is to use such a formalism to handle representing and reasoning about multiple disease in multimorbidity as discussed in Section 2.

References

[1] L. Cardelli. A semantics of multiple inheritance. *Information and Computation*, 76:138–164, 1988.

[2] L. Getoor, N. Friedman, D. Koller, A. Pfeffer, and B. Taskar. Probabilistic relational models. In *Introduction to statistical relational learning*. The MIT Press, 2009.

[3] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.

[4] D. Koller and A. Pfeffer. Object-oriented bayesian networks. In *Proceedings of Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, pages 302–313, 1997.

[5] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[6] L. Torti, P-H. Wuillemin, and C. Gonzales. Reinforcing the object-oriented aspect of probabilistic relational models. In *Proceedings of Probabilistic Graphical Models 2010*, pages 273–280, 2010.