

Monoidal Kleisli Category as a Background for Information Transformers Theory

Peter V. Golubtsov

*Moscow State Lomonosov University
Department of Physics, Moscow State University
119899, Moscow, Russia
E-mail: P_V_G@mail.ru
Received May 13, 2002*

CONTENTS

Abstract—We consider any uniform class of information transformers (ITs) as a family of morphisms of a monoidal category that contains a subcategory (of deterministic ITs) with finite products and satisfies certain set of axioms. Besides, many IT-categories can be constructed as Kleisli categories. The ingredients for this construction are: a base category (of deterministic ITs); a functor, producing objects of “distributions”; a natural transformation, representing “independent product of distributions”. The paper also generalizes Bayesian approach to decision-making problems and studies informativeness of ITs. It shows that classes of equivalent ITs form a partially ordered bounded Abelian monoid. Several examples of concrete IT-categories are examined.

1. INTRODUCTION

Currently the growing interest is attracted to various mathematical ways of describing uncertainty, most of them being different from the probabilistic one, (e.g., based on the apparatus of fuzzy sets). For adequate theoretical study of the corresponding “nonstochastic” systems of information transforming and, in particular, for the study of important notions, such as sufficiency, informativeness, etc., we need to develop an approach general enough to describe different classes of information transforming systems in a uniform way.

It is convenient to consider different systems that take place in information acquiring and processing as particular cases of so-called *information transformers* (ITs). Besides, it is useful to work with families of ITs in which certain operations, e.g., *sequential* and *parallel compositions* are defined.

It was noticed fairly long ago [1, 2, 3, 4, 5], that the adequate algebraic structure for describing information transformers (initially for the study of statistical experiments) is the structure of *category* [6, 7, 8, 9].

Analysis of general properties for the classes of linear, multivalued, and fuzzy information transformers, studied in [5, 10, 11, 12, 13, 14, 15, 16], allowed to extract general features shared by all these classes. Namely, each of these classes can be considered as a family of morphisms in an appropriate category, where the composition of information transformers corresponds to their “consecutive application.” Each category of ITs (or IT-category) contains a subcategory (of so called, *deterministic* ITs) that has products. Moreover, the operation of morphism product is extended in a “coherent way” to the whole category of ITs.

The work [17] undertook an attempt to formulate a set of “elementary” axioms for a category of ITs, which would be sufficient for an abstract expression of the basic concepts of the theory of information transformers and for study of informativeness, decision problems, etc. This paper proposes another, significantly more compact axiomatic for a category of ITs. According to this axiomatic a category of ITs is defined in effect as a *monoidal* category [6, 8], containing a subcategory (of *deterministic* ITs) with finite products.

Among the basic concepts connected to information transformers there is one that plays an important role in the uniform construction of a wide spectrum of IT-categories — the concept of *distribution*. Indeed, fairly often an IT $\alpha: \mathcal{A} \rightarrow \mathcal{B}$ can be represented by a mapping from \mathcal{A} to the “space of distributions” on \mathcal{B} (see, e.g., [13, 14]). For example, a probabilistic transition distribution (an IT in the category of stochastic ITs) can be represented by a certain measurable mapping from \mathcal{A} to the space of distributions on \mathcal{B} . This observation suggests to construct a category of ITs as a *Kleisli* category [6, 18, 19], arising from the following components: an obvious category of deterministic ITs; a functor that takes an object \mathcal{A} to the object of “distributions” on \mathcal{A} ; and a natural transformation of functors, describing an “independent product of distributions”.

It appears that rather general axiomatic theory, obtained this way, makes it possible to express in terms of IT-categories basic concepts for information transformers and to derive their main properties.

Of course, the most developed theory of uncertainty is probability theory (and statistics, based on probability). Certainly, mathematical statistics accumulated a rich conceptual experience. It introduced and deeply investigated such notions as joint and conditional distributions, independence, sufficiency, and others.

At the same time, it appears that all these concepts have very abstract meaning and hence, they can be treated in terms of alternative (i.e., not probabilistic) approaches to the description of uncertainty. In fact, the basic notions of probability theory and statistics, as well as the methodology and results, are easily extended to other theories dealing with uncertainty. In [12, 14] it is shown that a rather substantive decision theory may be constructed even on the very moderate basis of multivalued or fuzzy maps.

The approach developed in this paper allows to express easily in terms of IT-categories such concepts as distribution, joint and conditional distributions, independence, and others. It is shown that on the basis of these concepts it is possible to formulate fairly general statement of decision-making problem with a prior information, which generalizes the Bayesian approach in the theory of statistical decisions. Moreover, the *Bayesian* principle, derived below, like its statistical prototype [20], reduces the problem of optimal decision strategy construction to a significantly simpler problem of finding optimal decision for a posterior distribution.

Among the most important concepts in categories of ITs is the concept of (relative) *informativeness* of information transformers. There are two different approaches to the concept of informativeness.

One of these approaches is based on analyzing the “relative positions” of information transformers in the corresponding mathematical structure. Roughly speaking, one information transformer is regarded as more informative than another one if with the aid of an additional information transformer the former one can be “transformed” to an IT, which is similar to (or more “accurate” than) the latter one. In fact, this means that all the information that can be obtained from the latter information transformer can be extracted from the former one as well.

The other approach to informativeness is based on treating information transformers as data sources for decision-making problems. Here, one information transformer is said to be semantically more informative than another if it provides better quality of decision making. Obviously, the notion of semantical informativeness depends on the class of decision-making problems under consideration.

In the classical researches of Blackwell [21, 22], the correspondence between informativeness (Blackwell sufficiency) and semantical informativeness (Blackwell informativeness) were investigated in a statistical context. These studies were extended by Morse, Sacsteder, and Chentsov [1, 2, 4, 3], who applied the category theory techniques to their studies of statistical systems.

It is interesting, that under very general conditions the relations of informativeness and semantical informativeness (with respect to a certain class of decision-making problems) coincide. Moreover, in some categories of ITs it is possible to point out *one* special decision problem, such that the resulting semantical informativeness coincides with informativeness.

Analysis of classes of equivalent (with respect to informativeness) information transformers shows that they form a partially ordered Abelian monoid with the smallest (also neutral) and the largest elements.

One of the objectives of this paper is to show that the basic constructions and propositions of probability theory and statistics playing the fundamental role in decision-making problems have meaningful counterparts in terms of IT-categories. Furthermore, some definitions and propositions (for example, the notion of conditional distribution and the Bayesian principle) in terms of IT-categories often have more transparent meanings. This provides an opportunity to look at the well known results from a different angle. What is even more significant, it makes it possible to apply the methodology of statistical decision-making in an alternative (not probabilistic) context.

Approaches, proposed in this work may provide a background for construction and study of new classes of ITs, in particular, dynamical nondeterministic ITs, which may provide an adequate description for information flows and information interactions evolving in time. Besides, a uniform approach to problems of information transformations may be useful for better understanding of information processes that take place in complex artificial and natural systems.

2. CATEGORIES OF INFORMATION TRANSFORMERS

2.1. Common structure of classes of information transformers

It is natural to assume that for any information transformer a there are defined a couple of spaces: \mathcal{A} and \mathcal{B} , the space of “inputs” (or input signals) and the space of “outputs” (results of measurement, transformation, processing, etc.). We will say that a “acts” from \mathcal{A} to \mathcal{B} and denote this as $a: \mathcal{A} \rightarrow \mathcal{B}$. It is important to note that typically an information transformer not only transforms signals, but also introduces some “noise”. In this case it is *nondeterministic* and cannot be represented just by a mapping from \mathcal{A} to \mathcal{B} .

It is natural to study information transformers of similar type by aggregating them into families endowed by a fairly rich algebraic structure [5, 10]. Specifically, it is natural to assume that families of ITs poses the following properties:

- (a) If $a: \mathcal{A} \rightarrow \mathcal{B}$ and $b: \mathcal{B} \rightarrow \mathcal{C}$ are two ITs, then their *composition* $b \circ a: \mathcal{A} \rightarrow \mathcal{C}$ is defined.
- (b) This operation of composition is *associative*.
- (c) There are certain *neutral* elements in these families, i.e., ITs that do not introduce any alterations. Namely, for any space \mathcal{B} there exist a corresponding IT $i_{\mathcal{B}}: \mathcal{B} \rightarrow \mathcal{B}$ such that $i_{\mathcal{B}} \circ a = a$ and $b \circ i_{\mathcal{B}} = b$.

Algebraic structures of this type are called *categories* [6, 8].

Furthermore, we will assume, that to every pair of information transformers, acting from the same space D to spaces \mathcal{A} and \mathcal{B} respectively, there corresponds a certain IT $a * b$ (called *product* of a and b) from D to $\mathcal{A} \times \mathcal{B}$. This IT in a certain sense “represents” both ITs a and b simultaneously. Specifically, ITs a and b can be “extracted” from $a * b$ by means of projections $\pi_{\mathcal{A}, \mathcal{B}}$ and $\nu_{\mathcal{A}, \mathcal{B}}$ from $\mathcal{A} \times \mathcal{B}$ to \mathcal{A} and \mathcal{B} , respectively, i.e., $\pi_{\mathcal{A}, \mathcal{B}} \circ (a * b) = a$, $\nu_{\mathcal{A}, \mathcal{B}} \circ (a * b) = b$. Note, that typically, an IT c such that $\pi_{\mathcal{A}, \mathcal{B}} \circ c = a$, $\nu_{\mathcal{A}, \mathcal{B}} \circ c = b$ is not unique, i.e., a category of ITs does not have products (in category-theoretic sense [6, 7, 8, 9]). Thus, the notion of a category of ITs demands for an accurate formalization.

Analysis of classes of information transformers studied in [5, 10, 11, 12, 13, 14, 15, 16], gives grounds to consider these classes as categories that satisfy certain fairly general conditions.

2.2. Categories: basic concepts

Recall that a category (see, for example, [6, 7, 8, 9]) \mathbf{C} consists of a class of objects $\text{Ob}(\mathbf{C})$, a class of morphisms (or arrows) $\text{Ar}(\mathbf{C})$, and a composition operation \circ for morphisms, such that:

- (a) To any morphism a there corresponds a certain pair of objects \mathcal{A} and \mathcal{B} (the source and the target of a), which is denoted $a: \mathcal{A} \rightarrow \mathcal{B}$.
- (b) To every pair of morphisms $a: \mathcal{A} \rightarrow \mathcal{B}$ and $b: \mathcal{B} \rightarrow \mathcal{C}$ their *composition* $b \circ a: \mathcal{A} \rightarrow \mathcal{C}$ is defined.

Moreover, the following axioms hold:

(c) The composition is *associative*:

$$c \circ (b \circ a) = (c \circ b) \circ a.$$

(d) To every object \mathcal{R} there corresponds an (*identity*) morphism $i_{\mathcal{R}}: \mathcal{R} \rightarrow \mathcal{R}$, so that

$$\forall a: \mathcal{A} \rightarrow \mathcal{B} \quad a \circ i_{\mathcal{A}} = a = i_{\mathcal{B}} \circ a.$$

A morphism $a: \mathcal{A} \rightarrow \mathcal{B}$ is called *isomorphism* if there exists a morphism $b: \mathcal{B} \rightarrow \mathcal{A}$ such that $a \circ b = i_{\mathcal{B}}$ and $b \circ a = i_{\mathcal{A}}$. In this case objects \mathcal{A} and \mathcal{B} are called *isomorphic*.

Morphisms $a: \mathcal{D} \rightarrow \mathcal{A}$ and $b: \mathcal{D} \rightarrow \mathcal{B}$ are called *isomorphic* if there exists an isomorphism $c: \mathcal{A} \rightarrow \mathcal{B}$ such that $c \circ a = b$.

An object \mathcal{Z} is called *terminal* object if for any object \mathcal{A} there exists a unique morphism from \mathcal{A} to \mathcal{Z} , which is denoted $z_{\mathcal{A}}: \mathcal{A} \rightarrow \mathcal{Z}$ in what follows.

A category \mathbf{D} is called a *subcategory* of a category \mathbf{C} if $\text{Ob}(\mathbf{D}) \subseteq \text{Ob}(\mathbf{C})$, $\text{Ar}(\mathbf{D}) \subseteq \text{Ar}(\mathbf{C})$, and morphism composition in \mathbf{D} coincide with their composition in \mathbf{C} .

It is said that a category has (pairwise) products if for every pair of objects \mathcal{A} and \mathcal{B} there exists their *product*, that is, an object $\mathcal{A} \times \mathcal{B}$ and a pair of morphisms $\pi_{\mathcal{A},\mathcal{B}}: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{A}$ and $\nu_{\mathcal{A},\mathcal{B}}: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{B}$, called projections, such that for any object \mathcal{D} and for any pair of morphisms $a: \mathcal{D} \rightarrow \mathcal{A}$ and $b: \mathcal{D} \rightarrow \mathcal{B}$ there exists a unique morphism $c: \mathcal{D} \rightarrow \mathcal{A} \times \mathcal{B}$, satisfying the following conditions:

$$\pi_{\mathcal{A},\mathcal{B}} \circ c = a, \quad \nu_{\mathcal{A},\mathcal{B}} \circ c = b. \quad (1)$$

We call such morphism c the *product of morphisms* a and b and denote it $a * b$.

It is easily seen that existence of products in a category implies the following equality:

$$(a * b) \circ d = (a \circ d) * (b \circ d). \quad (2)$$

In a category with products, for two arbitrary morphisms $a: \mathcal{A} \rightarrow \mathcal{C}$ and $b: \mathcal{B} \rightarrow \mathcal{D}$ one can define the morphism $a \times b$:

$$a \times b: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{C} \times \mathcal{D}, \quad a \times b \stackrel{\text{def}}{=} (a \circ \pi_{\mathcal{A},\mathcal{B}}) * (b \circ \nu_{\mathcal{A},\mathcal{B}}). \quad (3)$$

This definition and (1) obviously imply that the morphism $c = a \times b$ satisfy the following conditions:

$$\pi_{\mathcal{C},\mathcal{D}} \circ c = a \circ \pi_{\mathcal{A},\mathcal{B}}, \quad \nu_{\mathcal{C},\mathcal{D}} \circ c = b \circ \nu_{\mathcal{A},\mathcal{B}} \quad (4)$$

Moreover, $c = a \times b$ is the only morphism satisfying conditions (4).

It is also easily seen that (2) and (3) imply the following equality:

$$(a \times b) \circ (c * d) = (a \circ c) * (b \circ d). \quad (5)$$

Suppose $\mathcal{A} \times \mathcal{B}$ and $\mathcal{B} \times \mathcal{A}$ are two products of objects \mathcal{A} and \mathcal{B} taken in different order. By the properties of products, the objects $\mathcal{A} \times \mathcal{B}$ and $\mathcal{B} \times \mathcal{A}$ are isomorphic and the natural isomorphism is

$$\sigma_{\mathcal{A},\mathcal{B}}: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{B} \times \mathcal{A}, \quad \sigma_{\mathcal{A},\mathcal{B}} \stackrel{\text{def}}{=} \nu_{\mathcal{A},\mathcal{B}} * \pi_{\mathcal{A},\mathcal{B}}. \quad (6)$$

Moreover, for any object \mathcal{D} and for any morphisms $a: \mathcal{D} \rightarrow \mathcal{A}$ and $b: \mathcal{D} \rightarrow \mathcal{B}$, the morphisms $a * b$ and $b * a$ are isomorphic, that is,

$$\sigma_{\mathcal{A},\mathcal{B}} \circ (a * b) = b * a. \quad (7)$$

Similarly, by the properties of products, the objects $(\mathcal{A} \times \mathcal{B}) \times \mathcal{C}$ and $\mathcal{A} \times (\mathcal{B} \times \mathcal{C})$ are isomorphic. Let

$$\alpha_{\mathcal{A},\mathcal{B},\mathcal{C}}: (\mathcal{A} \times \mathcal{B}) \times \mathcal{C} \rightarrow \mathcal{A} \times (\mathcal{B} \times \mathcal{C})$$

be the corresponding natural isomorphism. Its “explicit” form is:

$$\alpha_{\mathcal{A},\mathcal{B},\mathcal{C}} \stackrel{\text{def}}{=} (\pi_{\mathcal{A},\mathcal{B}} \circ \pi_{\mathcal{A} \times \mathcal{B},\mathcal{C}}) * \left((\nu_{\mathcal{A},\mathcal{B}} \circ \pi_{\mathcal{A} \times \mathcal{B},\mathcal{C}}) * \nu_{\mathcal{A} \times \mathcal{B},\mathcal{C}} \right). \quad (8)$$

Then for any object \mathcal{D} and for any morphisms $a: \mathcal{D} \rightarrow \mathcal{A}$, $b: \mathcal{D} \rightarrow \mathcal{B}$, and $c: \mathcal{D} \rightarrow \mathcal{C}$ we have

$$\alpha_{\mathcal{A},\mathcal{B},\mathcal{C}} \circ ((a * b) * c) = a * (b * c). \quad (9)$$

2.3. Elementary axioms for categories of information transformers

In this subsection we set forward the main properties of categories of ITs. All the following study will rely exactly on these properties.

In [5, 10, 11, 13, 16] it is shown (see also examples in section 6 below) that classes of information transformers can be considered as morphisms in certain categories. As a rule, such categories do not have products, which is a peculiar expression of nondeterministic nature of ITs in these categories. However, it turns out that deterministic information transformers, which are usually determined in a natural way in any category of ITs, form a subcategory with products. This point makes it possible to define a “product” of objects in a category of ITs. Moreover, it provides an axiomatic way to describe an extension of the product operation from the subcategory of deterministic ITs to the whole category of ITs.

Definition 1. We shall say that a category \mathbf{C} is a *category of information transformers* if the following axioms hold:

1. There is a fixed subcategory of *deterministic* ITs \mathbf{D} that contains all the objects of the category \mathbf{C} ($\text{Ob}(\mathbf{D}) = \text{Ob}(\mathbf{C})$).
2. The classes of *isomorphisms* in \mathbf{D} and in \mathbf{C} coincide, that is, all the isomorphisms in \mathbf{C} are deterministic.
3. The categories \mathbf{D} and \mathbf{C} have a common *terminal object* \mathcal{Z} .
4. The category \mathbf{D} has pairwise *products*.
5. There is a specified *extension of morphism product* from the subcategory \mathbf{D} to the whole category \mathbf{C} , that is, for any object \mathcal{D} and for any pair of morphisms $a: \mathcal{D} \rightarrow \mathcal{A}$ and $b: \mathcal{D} \rightarrow \mathcal{B}$ in \mathbf{C} there is certain information transformer $a * b: \mathcal{D} \rightarrow \mathcal{A} \times \mathcal{B}$ (which is also called a *product* of ITs a and b) such that

$$\pi_{\mathcal{A},\mathcal{B}} \circ (a * b) = a, \quad \nu_{\mathcal{A},\mathcal{B}} \circ (a * b) = b.$$

6. Let $a: \mathcal{A} \rightarrow \mathcal{C}$ and $b: \mathcal{B} \rightarrow \mathcal{D}$ are arbitrary ITs in \mathbf{C} , then the IT $a \times b$ defined by Eq. (3) satisfies Eq. (5):

$$(a \times b) \circ (c * d) = (a \circ c) * (b \circ d).$$

7. Equality (7) holds not only in \mathbf{D} but in \mathbf{C} as well, that is, *product* of information transformers is “*commutative up to isomorphism*.”
8. Equality (9) also holds in \mathbf{C} . In other words, *product* of information transformers is “*associative up to isomorphism*” too.

Now let us make several comments concerning the above definition.

We stress that in the description of the extension of morphism product from the category \mathbf{D} to \mathbf{C} (Axiom 5) we *do not require the uniqueness* of an IT $c: \mathcal{D} \rightarrow \mathcal{A} \times \mathcal{B}$ that satisfies conditions (1).

Nevertheless, it is easily verified, that the equations (4) are valid for $c = a \times b$ not only in the category \mathbf{D} , but in \mathbf{C} as well, that is,

$$\pi_{\mathcal{C},\mathcal{D}} \circ (a \times b) = a \circ \pi_{\mathcal{A},\mathcal{B}}, \quad \nu_{\mathcal{C},\mathcal{D}} \circ (a \times b) = b \circ \nu_{\mathcal{A},\mathcal{B}}.$$

However, the IT c that satisfy the equations (4) may be not unique. Note also that in the category \mathbf{C} Eq. (2) in general does not hold.

Further, note that Axiom 6 immediately implies

$$(a \times b) \circ (c \times d) = (a \circ c) \times (b \circ d).$$

Finally note that any category that has a terminal object and pairwise products can be considered as a category of ITs in which all information transformers are deterministic.

3. CATEGORY OF INFORMATION TRANSFORMERS AS A MONOIDAL CATEGORY

As we have already mentioned above in a category of ITs there are certain “meaningful” operations of product for objects and for morphisms. However, these operations are not product operations in category-theoretic sense. Nevertheless, every category of ITs is a *monoidal category* (see, e.g., [6, 8]).

First note, that every category \mathbf{D} with pairwise products and with terminal object \mathcal{Z} constitutes a monoidal category $\langle \mathbf{D}, \times, \mathcal{Z}, \alpha, \lambda, \rho \rangle$, where $\times: \mathbf{D} \times \mathbf{D} \rightarrow \mathbf{D}$ is the product functor and $\alpha_{\mathcal{A}, \mathcal{B}, \mathcal{C}}: (\mathcal{A} \times \mathcal{B}) \times \mathcal{C} \rightarrow \mathcal{A} \times (\mathcal{B} \times \mathcal{C})$, $\lambda_{\mathcal{A}}: \mathcal{Z} \times \mathcal{A} \rightarrow \mathcal{A}$, and $\rho_{\mathcal{A}}: \mathcal{A} \times \mathcal{Z} \rightarrow \mathcal{A}$ are the obvious natural equivalences. Besides, as a category with products, the category \mathbf{D} has a natural equivalence $\sigma, \sigma_{\mathcal{A}, \mathcal{B}}: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{B} \times \mathcal{A}$, which interchanges components in a product.

Definition 2. We will say that a category \mathbf{C} is a category of information transformers over a subcategory (of deterministic ITs) \mathbf{D} if the following three axioms hold.

Axiom 1. $\langle \mathbf{C}, \times, \mathcal{Z}, \alpha, \lambda, \rho \rangle$ is a monoidal category for a certain: functor $\times: \mathbf{C} \times \mathbf{C} \rightarrow \mathbf{C}$, object \mathcal{Z} , and natural equivalences α, λ and ρ .

We will refer to morphisms of the category \mathbf{C} as *information transformers*.

Axiom 2. The category \mathbf{C} has a subcategory \mathbf{D} , such that all the objects of \mathbf{C} are contained in \mathbf{D} , \mathcal{Z} is a terminal object in \mathbf{D} , and the functor \times is a product functor on \mathbf{D} .

Morphisms of the subcategory \mathbf{D} will be called *deterministic* information transformers.

Thus, the following properties hold in the subcategory \mathbf{D} :

- (a) There are natural transformations defined in \mathbf{D} , $\pi_{\mathcal{A}, \mathcal{B}}: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{A}$ and $\nu_{\mathcal{A}, \mathcal{B}}: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{B}$ that specify projections on components of a product.
- (b) For any deterministic ITs (morphisms in \mathbf{D}) $a: \mathcal{C} \rightarrow \mathcal{A}$ and $b: \mathcal{C} \rightarrow \mathcal{B}$ there exists a unique IT $c = a * b: \mathcal{C} \rightarrow \mathcal{A} \times \mathcal{B}$ for which $\pi_{\mathcal{A}, \mathcal{B}} \circ c = a$ and $\nu_{\mathcal{A}, \mathcal{B}} \circ c = b$;
- (c) \mathbf{D} is also a monoidal category with the natural equivalences α, λ and ρ explicitly expressed through π and ν , i.e.,

$$\begin{aligned} \lambda_{\mathcal{A}} &\stackrel{\text{def}}{=} \pi_{\mathcal{Z}, \mathcal{A}}, & \rho_{\mathcal{A}} &\stackrel{\text{def}}{=} \nu_{\mathcal{A}, \mathcal{Z}}, \\ \alpha_{\mathcal{A}, \mathcal{B}, \mathcal{C}} &\stackrel{\text{def}}{=} (\pi_{\mathcal{A}, \mathcal{B}} \circ \pi_{\mathcal{A} \times \mathcal{B}, \mathcal{C}}) * \left((\nu_{\mathcal{A}, \mathcal{B}} \circ \pi_{\mathcal{A} \times \mathcal{B}, \mathcal{C}}) * \nu_{\mathcal{A} \times \mathcal{B}, \mathcal{C}} \right). \end{aligned}$$

- (d) There is a natural equivalence of “object transposition” σ defined on \mathbf{D} :

$$\sigma_{\mathcal{A}, \mathcal{B}} \stackrel{\text{def}}{=} \nu_{\mathcal{A}, \mathcal{B}} * \pi_{\mathcal{A}, \mathcal{B}}: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{B} \times \mathcal{A}.$$

- (e) There is a “diagonal” natural transformation δ defined on \mathbf{D} :

$$\delta_{\mathcal{C}} \stackrel{\text{def}}{=} i_{\mathcal{C}} * i_{\mathcal{C}}: \mathcal{C} \rightarrow \mathcal{C} \times \mathcal{C}.$$

Note that with the help of the “diagonal” natural transformation the product of morphisms $a * b$ may be expressed through their “functorial product” $a \times b$, i.e., $a * b = (a \times b) \circ \delta_C$.

Axiom 3. Natural transformations π, ν and σ (in the category **D**) are natural transformations in the whole category **C** as well.

Let us stress here, that we do not require that δ is a natural transformation on the whole category **C**. Furthermore, typically, in many important examples of categories of ITs δ is not a natural transformation. Such categories do not have products in category-theoretic sense. However we can extend the product operation for morphisms from the subcategory **D** to **C**. Specifically, we define in **C**:

$$a * b \stackrel{\text{def}}{=} (a \times b) \circ \delta_C.$$

Theorem 1. *Definitions 1 and 2 for a category of information transformers are equivalent.*

4. IT-CATEGORY AS KLEISLI CATEGORY

4.1. Concept of distribution. Kleisli category

The two equivalent definitions presented above provide the minimal conceptual background for studying categories of ITs, e.g., for definition and analysis of informativeness, semantic informativeness, decision problems, etc. [1, 2, 3, 4, 5, 17, 13, 10]. However these definitions do not provide any tools for constructing categories of ITs on the basis of more elementary concepts. The concept of *distribution* is one of the most important and it plays a critical role in the uniform construction of a wide spectrum of IT-categories. Its importance is connected to the observation that in many important IT-categories an information transformer $a: \mathcal{A} \rightarrow \mathcal{B}$ may be represented by a morphisms from \mathcal{A} to the “object of distributions” over \mathcal{B} . For example, a probabilistic transition distribution (an IT in the category of stochastic ITs) may be represented by a certain measurable mapping \mathcal{A} to the space of distributions on \mathcal{B} .

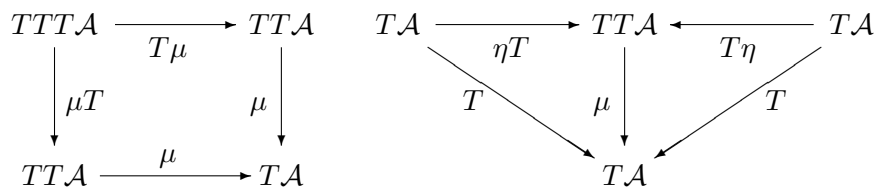
Thus, we will suppose that on some fixed “base” category **D** (category of deterministic ITs) there defined a functor T , which takes an object \mathcal{A} to the object $T\mathcal{A}$ of “distributions” on \mathcal{A} . Besides, we assume that there are two natural transformations connected to this functor: $\eta: I \rightarrow T$ and $\mu: TT \rightarrow T$. Informally, $\eta_A: \mathcal{A} \rightarrow T\mathcal{A}$ takes an element of \mathcal{A} to a “discrete distribution, concentrated on this element”, and $\mu_A: TT\mathcal{A} \rightarrow T\mathcal{A}$ “mixes” (averages) a distribution of distributions on \mathcal{A} , by transforming it to a certain distribution on \mathcal{A} . Besides, there are natural “coherence” conditions for η and μ :

$$\mu_A \circ T\mu_A = \mu_A \circ \mu_{T\mathcal{A}}$$

and

$$\mu_A \circ T\eta_A = i_{T\mathcal{A}} \quad \mu_A \circ \eta_{T\mathcal{A}} = i_{T\mathcal{A}}$$

that may be presented by the following commutative diagrams:



Commutativity of the square means that for any “third-order distribution” on \mathcal{A} (i.e. distribution on a collection of distributions on a family of distributions on \mathcal{A}) the result of “mixing” of distributions does not depend on the order of “mixing”. More precisely, the result of mixing over the “top” (third order, element of $TTT\mathcal{A}$) distribution first and mixing the resulting second-order distribution next should give the same result as for mixing over “intermediate” (second-order, elements of $TT\mathcal{A}$) distributions first and then mixing the

resulting second order distribution. Commutativity of the left triangle means that mixing of a second order distribution, concentrated in one element (which is itself a distribution on \mathcal{A}) gives this distribution. Finally, commutativity of the right triangle means if we take some distribution on \mathcal{A} , transform it to “the same” distribution of singletons and then mix the resulting second-order distribution, we will obtain the original distribution.

It is well known, that a collection $\langle T, \eta, \mu \rangle$, satisfying the two commutative diagrams above, is called a *triple* (monad) [6, 8, 19] on the category \mathbf{D} .

The concept of triple provides an elegant technique of constructing a category of ITs \mathbf{C} on the basis of the category of deterministic ITs, as a *Kleisli* category [6, 18, 19]. In this construction each morphisms $a: \mathcal{A} \rightarrow \mathcal{B}$ in the category \mathbf{C} is determined by a morphism $a': \mathcal{A} \rightarrow T\mathcal{B}$ of the category \mathbf{D} . The composition $a \circ b$ of ITs $a: \mathcal{A} \rightarrow \mathcal{B}$ and $b: \mathcal{B} \rightarrow \mathcal{C}$ in \mathbf{C} is represented by the morphism

$$(b \circ a)' \stackrel{\text{def}}{=} \mu_{\mathcal{C}} \circ T b' \circ a'$$

$$(b \circ a)' = \mathcal{A} \xrightarrow{a'} T\mathcal{B} \xrightarrow{T b'} T T\mathcal{C} \xrightarrow{\mu} T\mathcal{C}$$

in \mathbf{D} , and any deterministic IT $c: \mathcal{C} \rightarrow \mathcal{D}$ (in \mathbf{C}) are determined by the morphism

$$c' \stackrel{\text{def}}{=} \eta_{\mathcal{D}} \circ c$$

$$c' = \mathcal{C} \xrightarrow{c} \mathcal{D} \xrightarrow{\eta} T\mathcal{D}$$

in \mathbf{D} .

4.2. Independent distribution. Monoidal Kleisli category

The main factor in the construction of the category of ITs as a Kleisli category is equipping it with a structure of monoidal category. For this purpose we introduce a natural transformation $\gamma: \times T \rightarrow T \times$, $\gamma_{\mathcal{A}, \mathcal{B}}: T\mathcal{A} \times T\mathcal{B} \rightarrow T(\mathcal{A} \times \mathcal{B})$, which “takes” a pair of distributions to their “independent joint distribution” (see also [23]). Then the product $c = a * b$ of ITs $a: \mathcal{D} \rightarrow \mathcal{A}$ and $b: \mathcal{D} \rightarrow \mathcal{B}$ (in \mathbf{C}) is determined by the morphism

$$c' \stackrel{\text{def}}{=} \gamma_{\mathcal{A}, \mathcal{B}} \circ (a' * b')$$

$$\begin{array}{ccccc}
 & & T\mathcal{A} & & \\
 & a' \nearrow & \uparrow \pi T & & \\
 \mathcal{D} & \xrightarrow{a' * b'} & T\mathcal{A} \times T\mathcal{B} & \xrightarrow{\gamma} & T(\mathcal{A} \times \mathcal{B}) \\
 & b' \searrow & \downarrow \nu T & & \\
 & & T\mathcal{B} & &
 \end{array}$$

in \mathbf{D} . Note, that $a' * b'$ here exists and is uniquely defined since \mathbf{D} is a category with products.

Theorem 2. *Suppose that \mathbf{D} is a category with pairwise products and with terminal object \mathcal{Z} ; π, ν, α, σ are the corresponding natural transformations, and $\langle T, \eta, \mu \rangle$ is a triple on \mathbf{D} with $\eta_{\mathcal{B}}$ monomorphic for every \mathcal{B} . Then the generated Kleisli category \mathbf{C} , equipped with a natural transformation γ , is a category of information transformers if and only if the following compatibility conditions of γ with the natural transformations $\pi, \nu, \alpha, \sigma, \eta$, and μ hold:*

π - γ and ν - γ conditions:

$$T\pi_{A,B} \circ \gamma_{A,B} = \pi_{TA,TB} \quad T\nu_{A,B} \circ \gamma_{A,B} = \nu_{TA,TB}$$

σ - γ condition:

$$T\sigma_{A,B} \circ \gamma_{A,B} = \gamma_{B,A} \circ \sigma_{TA,TB}$$

α - γ condition:

$$T\alpha_{A,B,C} \circ \gamma_{A \times B, C} \circ (\gamma_{A,B} \times i_{TC}) = \gamma_{A, B \times C} \circ (i_{TA} \times \gamma_{B,C}) \circ \alpha_{TA, TB, TC}$$

μ - γ condition:

$$\mu_{A \times B} \circ T\gamma_{A,B} \circ \gamma_{TA, TB} = \gamma_{A,B} \circ (\mu_A \times \mu_B)$$

η - γ condition:

$$\gamma_{A,B} \circ (\eta_A \times \eta_B) = \eta_{A \times B}$$

Thus, construction of a categories of ITs is, in effect, reduced to selection of a base category \mathbf{D} , a functor $T: \mathbf{D} \rightarrow \mathbf{D}$, and a natural transformation $\gamma: \times T \rightarrow T \times$.

All these conditions have rather transparent meaning that we will try to comment below.

For better understanding we also provide the corresponding commutative diagrams in which we omit the obvious indices for the sake of readability:

π - γ and ν - γ conditions. Marginal distributions extracted from independent joint distribution coincide with the original distributions:

$$\begin{array}{ccc} & TA & \\ \pi T \nearrow & & \nwarrow T\pi \\ TA \times TB & \xrightarrow{\gamma} & T(A \times B) \\ \nu T \searrow & & \nearrow T\nu \\ & TB & \end{array}$$

σ - γ condition. Transposition of components of an independent joint distribution leads to the corresponding transformation of the joint distribution, i.e., Independent joint distribution is “invariant” with respect to transposition of its components. More precisely, we can say that the independent distribution morphism for transposed components $\gamma_{B,A}: TB \times TA \rightarrow T(B \times A)$ is naturally isomorphic to the original morphism $\gamma_{A,B}: TA \times TB \rightarrow T(A \times B)$. The corresponding isomorphism (of morphisms) is provided by the pair $\langle \sigma_{TA, TB}, T\sigma_{A,B} \rangle$:

$$\begin{array}{ccc} TA \times TB & \xrightarrow{\gamma} & T(A \times B) \\ \downarrow \sigma T & & T\sigma \downarrow \\ TB \times TA & \xrightarrow{\gamma} & T(B \times A) \end{array}$$

α - γ condition: Independent joint distribution for three components is “naturally invariant” with respect to the order of parentheses. More precisely, the morphisms

$$\gamma_{A, B \times C} \circ (i_{TA} \times \gamma_{B,C}): TA \times (TB \times TC) \rightarrow T(A \times (B \times C))$$

and

$$\gamma_{\mathcal{A} \times \mathcal{B}, \mathcal{C}} \circ (\gamma_{\mathcal{A}, \mathcal{B}} \times i_{TC}): (T\mathcal{A} \times T\mathcal{B}) \times TC \rightarrow T((\mathcal{A} \times \mathcal{B}) \times \mathcal{C})$$

(that take independent joint distributions for three components with different order of parentheses) are naturally isomorphic via $\langle \alpha_{T\mathcal{A}, T\mathcal{B}, TC}, T\alpha_{\mathcal{A}, \mathcal{B}, \mathcal{C}} \rangle$:

$$\begin{array}{ccccc} (T\mathcal{A} \times T\mathcal{B}) \times TC & \xrightarrow{\gamma \times T} & T(\mathcal{A} \times \mathcal{B}) \times TC & \xrightarrow{\gamma} & T((\mathcal{A} \times \mathcal{B}) \times \mathcal{C}) \\ \downarrow \alpha T & & & & \downarrow T\alpha \\ T\mathcal{A} \times (T\mathcal{B} \times TC) & \xrightarrow{T \times \gamma} & T\mathcal{A} \times T(\mathcal{B} \times \mathcal{C}) & \xrightarrow{\gamma} & T(\mathcal{A} \times (\mathcal{B} \times \mathcal{C})) \end{array}$$

μ - γ condition. Independent joint distribution for results of mixing of two second-order distributions may also be obtained by mixing the corresponding second-order independent distributions:

$$\begin{array}{ccccc} TTA \times TT\mathcal{B} & \xrightarrow{\gamma T} & T(T\mathcal{A} \times T\mathcal{B}) & \xrightarrow{T\gamma} & TT(\mathcal{A} \times \mathcal{B}) \\ \downarrow \mu \times \mu & & & & \downarrow \mu \\ T\mathcal{A} \times T\mathcal{B} & \xrightarrow{\gamma} & & & T(\mathcal{A} \times \mathcal{B}) \end{array}$$

η - γ condition: Independent joint distribution for two “singleton” distributions is just the corresponding “singleton” distribution on a product space:

$$\begin{array}{ccc} & \mathcal{A} \times \mathcal{B} & \\ \eta \times \eta \swarrow & & \searrow \eta \\ T\mathcal{A} \times T\mathcal{B} & \xrightarrow{\gamma} & T(\mathcal{A} \times \mathcal{B}) \end{array}$$

5. INFORMATIVENESS OF INFORMATION TRANSFORMERS

5.1. Accuracy relation

In order to define informativeness relation we will need to introduce first the following auxiliary notion.

Definition 3. We will say that \triangleright is an *accuracy* relation on an IT-category \mathbf{C} if for any pair of objects \mathcal{A} and \mathcal{B} in \mathbf{C} the set $\mathbf{C}(\mathcal{A}, \mathcal{B})$ of all ITs from \mathcal{A} to \mathcal{B} is equipped with a partial order \triangleright that satisfies the following *monotonicity* conditions:

$$\begin{aligned} a \triangleright a', b \triangleright b' &\implies a \circ b \triangleright a' \circ b', \\ a \triangleright a', b \triangleright b' &\implies a * b \triangleright a' * b'. \end{aligned}$$

Thus, the composition and the product are monotonous with respect to the partial order \triangleright . For a pair of ITs $a, b \in \mathbf{C}(\mathcal{A}, \mathcal{B})$ we shall say that a is more *accurate* than b whenever $a \triangleright b$.

It obviously follows from the very definition of the operation \times (3) and from the monotonicity conditions that the operation \times is monotone as well:

$$a \triangleright a', b \triangleright b' \implies a \times b \triangleright a' \times b'.$$

It is clear that for any IT-category there exists at least a “trivial variant” of the partial order \triangleright , namely, one can choose an equality relation for \triangleright , that is, one can put $a \triangleright b \stackrel{\text{def}}{\iff} a = b$. However, many categories of ITs (for example, multivalued and fuzzy ITs) provide a “natural” choice of the accuracy relation, which is different from the equality relation.

5.2. Definition of informativeness relation

Suppose $a: \mathcal{D} \rightarrow \mathcal{A}$ and $b: \mathcal{D} \rightarrow \mathcal{B}$ are two information transformers with a common source \mathcal{D} . Assume that there exists an IT $c: \mathcal{A} \rightarrow \mathcal{B}$ such that $c \circ a = b$. Then any information that can be obtained from b can be obtained from a as well (by attaching the IT c next to a). Thus, it is natural to consider the information transformer a as being more informative than the IT b and also more informative than any IT less accurate than b .

Now we give the formal definition of the informativeness relation in the category of information transformers.

Definition 4. We shall say that an information transformer a is more *informative* (better) than b if there exists an information transformer c such that $c \circ a \triangleright b$, that is,

$$a \succcurlyeq b \stackrel{\text{def}}{\iff} \exists c \ c \circ a \triangleright b.$$

It is easily verified that the informativeness relation \succcurlyeq is a preorder on the class of information transformers in \mathbf{C} . This preorder \succcurlyeq induces an equivalence relation \sim in the following way:

$$a \sim b \stackrel{\text{def}}{\iff} a \succcurlyeq b \ \& \ b \succcurlyeq a.$$

Obviously, the relation “more informative” extends the relation “more accurate,” that is,

$$a \triangleright b \implies a \succcurlyeq b.$$

5.3. Main properties of informativeness

It can be easily verified that the informativeness relation \succcurlyeq satisfies the following natural properties.

Lemma 1. Consider all information transformers with a fixed source \mathcal{D} .

- (a) The identity information transformer $i_{\mathcal{D}}$ is the most informative and the terminal information transformer $z_{\mathcal{D}}$ is the least informative:

$$\forall a \quad i_{\mathcal{D}} \succcurlyeq a \succcurlyeq z_{\mathcal{D}}.$$

- (b) Any information transformer $a: \mathcal{D} \rightarrow \mathcal{B} \times \mathcal{C}$ is more informative than its parts $\pi_{\mathcal{B},\mathcal{C}} \circ a$ and $\nu_{\mathcal{B},\mathcal{C}} \circ a$.
(c) The product $a * b$ is more informative than its components

$$a * b \succcurlyeq a, b.$$

Furthermore, the informativeness relation is compatible with the composition and the product operations.

Lemma 2. (a) If $a \succcurlyeq b$, then $a \circ c \succcurlyeq b \circ c$.

(b) If $a \succcurlyeq b$ and $c \succcurlyeq e$, then $a * c \succcurlyeq b * e$.

5.4. Structure of the family of informativeness equivalence classes

Let a be some information transformer. We shall denote by $[a]$ the equivalence (with respect to informativeness) class of a . We shall also use boldface for equivalence classes, that is, $a \in \mathbf{a}$ is equivalent to $\mathbf{a} = [a]$.

Theorem 3. Let $\mathfrak{J}(\mathcal{D})$ be the family of informativeness equivalence classes for the class of all information transformers with a fixed domain \mathcal{D} . The family $\mathfrak{J}(\mathcal{D})$ forms a partial ordered Abelian monoid $\langle \mathfrak{J}(\mathcal{D}), \succcurlyeq, *, \mathbf{0} \rangle$ with the smallest element $\mathbf{0}$ and the largest element $\mathbf{1}$, where

$$[a] \succcurlyeq [b] \stackrel{\text{def}}{\iff} a \succcurlyeq b, \quad [a] * [b] \stackrel{\text{def}}{=} [a * b], \quad \mathbf{0} \stackrel{\text{def}}{=} [z_{\mathcal{D}}], \quad \mathbf{1} \stackrel{\text{def}}{=} [i_{\mathcal{D}}].$$

Moreover, the following properties hold:

- (a) $\mathbf{0} * \mathbf{a} = \mathbf{a}$,
- (b) $\mathbf{1} * \mathbf{a} = \mathbf{1}$,
- (c) $\mathbf{0} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1}$,
- (d) $\mathbf{a} * \mathbf{b} \succcurlyeq \mathbf{a}, \mathbf{b}$,
- (e) $(\mathbf{a} \succcurlyeq \mathbf{b}) \ \& \ (\mathbf{c} \succcurlyeq \mathbf{e}) \implies \mathbf{a} * \mathbf{c} \succcurlyeq \mathbf{b} * \mathbf{e}$.

6. INFORMATIVENESS AND SYNTHESIS OF OPTIMAL INFORMATION TRANSFORMERS

In this section, we consider an alternative (with respect to the above) approach to informativeness comparison. This approach is based on treating information transformers as data sources for decision-making problems.

6.1. Decision-making problems in categories of ITs

Results of observations, obtained on real sources of information (e.g. indirect measurements) are as a rule unsuitable for straightforward interpretation. Typically it is assumed that observations suitable for interpretation are those into a certain object \mathcal{U} which in what follows will be called object of *interpretations* or object of *decisions*.

By an *interpretable* information transformer for signals from an object \mathcal{D} we mean any information transformer $a: \mathcal{D} \rightarrow \mathcal{U}$.

It is usually thought that some interpretable information transformers are more suitable for interpretation (of obtained results) than others. Namely, on a set $\mathbf{C}(\mathcal{D}, \mathcal{U})$ of information transformers from \mathcal{D} to \mathcal{U} , one defines some preorder relation \gg , which specifies the relative quality of various interpretable information transformers. Typically the relation \gg is predetermined by the specific formulation of a problem of optimal information transformer synthesis (that is, decision-making problem).

We shall say that an abstract *decision-making problem* is determined by a triple $\langle \mathcal{D}, \mathcal{U}, \gg \rangle$, where \mathcal{D} is an object of studied (input) signals, \mathcal{U} is an object of decisions (or interpretations), and \gg is a preorder on the set $\mathbf{C}(\mathcal{D}, \mathcal{U})$.

We shall call a preorder \gg *monotone* if for any $a, b \in \mathbf{C}(\mathcal{D}, \mathcal{U})$

$$a \triangleright b \implies a \gg b,$$

that is, more accurate IT provides better quality of interpretation.

For a given information transformer $a: \mathcal{D} \rightarrow \mathcal{A}$ we shall also say that an IT b *reduces* a to an interpretable information transformer if $b \circ a: \mathcal{D} \rightarrow \mathcal{U}$, that is, if $b: \mathcal{A} \rightarrow \mathcal{U}$. Such an information transformer b will be called a *decision strategy*.

The set of all interpretable information transformers obtainable on the basis of $a: \mathcal{D} \rightarrow \mathcal{A}$ will be denoted $\mathbf{U}_a \subseteq \mathbf{C}(\mathcal{D}, \mathcal{U})$:

$$\mathbf{U}_a \stackrel{\text{def}}{=} \{b \circ a \mid b: \mathcal{A} \rightarrow \mathcal{U}\}.$$

We shall call a decision strategy $r: \mathcal{A} \rightarrow \mathcal{U}$ *optimal* (for the IT a with respect to the problem $\langle \mathcal{D}, \mathcal{U}, \gg \rangle$) if the IT $r \circ a$ is a maximal element in \mathbf{U}_a with respect to \gg . Thus, a decision-making problem for a given information transformer a is stated as the problem of constructing optimal decision strategies.

6.2. Semantical informativeness

The relation \gg induces a preorder relation \sqsubseteq on a class of information transformers operating from \mathcal{D} in the following way.

Assume that a and b are information transformers with the source \mathcal{D} , that is, $a: \mathcal{D} \rightarrow \mathcal{A}$, $b: \mathcal{D} \rightarrow \mathcal{B}$. By definition, put

$$a \sqsubseteq b \stackrel{\text{def}}{\iff} \forall b': \mathcal{B} \rightarrow \mathcal{U} \exists a': \mathcal{A} \rightarrow \mathcal{U} \quad a' \circ a \gg b' \circ b.$$

In other words, $a \sqsubseteq b$ if for every interpretable information transformer d derived from b there exists an interpretable information transformer c derived from a such that $c \gg d$, that is,

$$a \sqsubseteq b \iff \forall d \in \mathbf{U}_b \exists c \in \mathbf{U}_a \quad c \gg d.$$

It can easily be checked that the relation \sqsubseteq is a preorder relation.

It is natural to expect that if one information transformer is more informative than the other, then the former will be better than the latter in any context. In other words, for any preorder \gg on the set of interpretable information transformers the induced preorder \sqsubseteq is dominated by the informativeness relation \succ (that is, \sqsubseteq is weaker than \succ). The converse is also true.

Definition 5. We shall say that an information transformer a is *semantically more informative* than b if for any interpretation object \mathcal{U} and for any preorder \gg (on the set of interpretable information transformers) $a \sqsubseteq b$ for the induced preorder \sqsubseteq .

The following theorem is in some sense a “completeness” theorem, which establishes a relation between “structure” (b can be “derived” from a) and “semantics” (a is uniformly better than b in decision-making problems).

Theorem 4. For any information transformers a and b with a common source \mathcal{D} , information transformer a is more informative than b if and only if a is semantically more informative than b .

Let us remark that the above proof relies heavily on the extreme extent of the class of decision problems involved. This makes it possible to select for any given pair of ITs a, b an appropriate decision-making problem $\langle \mathcal{D}, \mathcal{U}_b, \gg_b \rangle$ in which the interpretation object \mathcal{U}_b and the preorder \gg_b depend on the IT b . However, in some cases it is possible to point out a concrete (universal) decision-making problem such that

$$a \succ b \iff a \sqsubseteq b.$$

Theorem 5. Assume that for a given object \mathcal{D} there exists an object $\tilde{\mathcal{D}}$ such that for every information transformer acting from \mathcal{D} there exists an equivalent (with respect to informativeness) IT acting from $\tilde{\mathcal{D}}$, that is,

$$\forall \mathcal{B} \quad \forall b: \mathcal{D} \rightarrow \mathcal{B} \quad \exists b': \mathcal{D} \rightarrow \tilde{\mathcal{D}} \quad b \sim b'.$$

Let us choose the decision object $\mathcal{U} \stackrel{\text{def}}{=} \tilde{\mathcal{D}}$ and the preorder \gg , defined by

$$c \gg d \stackrel{\text{def}}{\iff} c \triangleright d.$$

Then $a \succ b$ if and only if $a \sqsubseteq b$.

Note that in general case an optimal decision strategy (if exists) can be nondeterministic. However, in many cases it is sufficient to search optimal strategies among deterministic ITs. Indeed, in some categories of information transformers the relation of “accuracy” satisfies the following condition: every IT is dominated by some deterministic IT, that is, for every IT there exists a more accurate deterministic IT.

Proposition 1. *Assume that $\langle \mathcal{D}, \mathcal{U}, \gg \rangle$ is a monotone decision-making problem in a category of ITs \mathbf{C} . Assume also that the following condition holds:*

$$\forall c \in \text{Ar}(\mathbf{C}) \quad \exists d \in \text{Ar}(\mathbf{D}) \quad d \triangleright c.$$

Then for any IT $a: \mathcal{D} \rightarrow \mathcal{R}$ and for any decision strategy $r: \mathcal{R} \rightarrow \mathcal{U}$ there exists a deterministic strategy $r_0: \mathcal{R} \rightarrow \mathcal{U}$ such that $r_0 \circ a \gg r \circ a$.

7. DECISION-MAKING PROBLEMS WITH A PRIOR INFORMATION

In this section we formulate in terms of categories of information transformers an analogy for the classical problem of optimal decision strategy construction for decision *problems with a prior information* (or information *a priori*). We also prove a counterpart of the *Bayesian principle* from the theory of statistical games [20, 24]. Like its statistical prototype it reduces the problem of constructing an optimal decision strategy to a much simpler problem of finding an optimal decision for a *posterior information* (or information *a posteriori*).

First we define in terms of categories of information transformers some necessary concepts, namely, concepts of distribution, conditional information transformer, decision problem with a prior information, and others.

7.1. Distributions in categories of ITs

We shall say that a *distribution* on an object \mathcal{A} (in some fixed category of ITs \mathbf{C}) is any IT $f: \mathcal{Z} \rightarrow \mathcal{A}$, where \mathcal{Z} is the terminal object in \mathbf{C} .

The concept of distribution corresponds to the general concept of an element of some object in a category, namely, a morphism from the terminal object (see, e.g., [9]).

Any distribution of the form $h: \mathcal{Z} \rightarrow \mathcal{A} \times \mathcal{B}$ will be called a *joint distribution* on \mathcal{A} and \mathcal{B} . The projections $\pi_{\mathcal{A}, \mathcal{B}}$ and $\nu_{\mathcal{A}, \mathcal{B}}$ on the components \mathcal{A} and \mathcal{B} respectively, “extract” *marginal distributions* f and g of the joint distribution h , that is,

$$\begin{aligned} f &= \pi_{\mathcal{A}, \mathcal{B}} \circ h: \mathcal{Z} \rightarrow \mathcal{A}, \\ g &= \nu_{\mathcal{A}, \mathcal{B}} \circ h: \mathcal{Z} \rightarrow \mathcal{B}. \end{aligned}$$

We say that the components of a joint distribution $h: \mathcal{Z} \rightarrow \mathcal{A} \times \mathcal{B}$ are *independent* whenever this joint distribution is completely determined by its marginal distributions, that is,

$$h = (\pi_{\mathcal{A}, \mathcal{B}} \circ h) * (\nu_{\mathcal{A}, \mathcal{B}} \circ h).$$

Let f be an arbitrary distribution on \mathcal{A} and let $a: \mathcal{A} \rightarrow \mathcal{B}$ be some information transformer. Then the distribution $g = a \circ f$ in some sense “contains an information about f .” This concept can be expressed precisely if one consider the joint distribution *generated* by the distribution f and the IT a :

$$h: \mathcal{Z} \rightarrow \mathcal{A} \times \mathcal{B}, \quad h = (i_{\mathcal{A}} * a) \circ f.$$

Note, that the marginal distributions for h coincide with f and g , respectively. Indeed,

$$\begin{aligned} \pi_{\mathcal{A}, \mathcal{B}} \circ h &= \pi_{\mathcal{A}, \mathcal{B}} \circ (i_{\mathcal{A}} * a) \circ f = i_{\mathcal{A}} \circ f = f, \\ \nu_{\mathcal{A}, \mathcal{B}} \circ h &= \nu_{\mathcal{A}, \mathcal{B}} \circ (i_{\mathcal{A}} * a) \circ f = a \circ f = g. \end{aligned}$$

Let h be a joint distribution on $\mathcal{A} \times \mathcal{B}$. We shall say that $a: \mathcal{A} \rightarrow \mathcal{B}$ is a *conditional IT* for h with respect to \mathcal{A} whenever h is generated by the marginal distribution $\pi_{\mathcal{A}, \mathcal{B}} \circ h$ and the IT a , that is,

$$h = (i_{\mathcal{A}} * a) \circ \pi_{\mathcal{A}, \mathcal{B}} \circ h.$$

Similarly, an IT $b: \mathcal{B} \rightarrow \mathcal{A}$ such that

$$h = (b * i_{\mathcal{B}}) \circ \nu_{\mathcal{A}, \mathcal{B}} \circ h$$

will be called a conditional IT for h with respect to \mathcal{B} .

7.2. Bayesian decision-making problems

Suppose that, like in Section 6, there are fixed two objects \mathcal{D} and \mathcal{U} in some category of ITs, namely, the object of signals and the object of decisions, respectively. In a decision-making problem with a prior distribution f on \mathcal{D} one fixes some preorder \gg_f on the set of joint distributions on $\mathcal{D} \times \mathcal{U}$ for which \mathcal{D} -marginal distribution coincides with f .

Informally, any joint distribution h on $\mathcal{D} \times \mathcal{U}$ of this kind can be considered as a joint distribution of a studied signal (with the distribution $f = \pi_{\mathcal{D}, \mathcal{U}} \circ h$ on \mathcal{D}) and a decision (with the distribution $g = \nu_{\mathcal{D}, \mathcal{U}} \circ h$ on \mathcal{U}). The preorder \gg_f determines how good is the “correlation” between studied signals and decisions.

Formally, an abstract *decision problem with a prior information* is determined by a quadruple $\langle \mathcal{D}, \mathcal{U}, f, \gg_f \rangle$, where \mathcal{D} is an object of studied signals, \mathcal{U} is an object of decisions (or interpretations), $f: \mathcal{Z} \rightarrow \mathcal{D}$ is a *prior* distribution (or distribution *a priori*), and \gg_f is a preorder on the set of ITs $h: \mathcal{Z} \rightarrow \mathcal{D} \times \mathcal{U}$ that satisfy the condition $\pi_{\mathcal{D}, \mathcal{U}} \circ h = f$.

Furthermore, suppose that there is a fixed IT $a: \mathcal{D} \rightarrow \mathcal{R}$ (which determines a measurement; \mathcal{R} can be called an object of *observations*). An IT $r: \mathcal{R} \rightarrow \mathcal{U}$ is called *optimal* (for the IT a with respect to \gg_f) if the distribution $(i * r \circ a) \circ f$ is a maximal element with respect to \gg_f . The set of all optimal information transformers is denoted $\text{Opt}_f(a \circ f)$.

Theorem 6. (Bayesian principle). *Let f be a given prior distribution on \mathcal{D} , let $a: \mathcal{D} \rightarrow \mathcal{R}$ be a fixed IT, and let $b: \mathcal{R} \rightarrow \mathcal{D}$ be a conditional information transformer for $(i * a) \circ f$ with respect to \mathcal{R} . Then the set of optimal ITs $r: \mathcal{R} \rightarrow \mathcal{U}$, namely, the set of optimal decision strategies for f over $a \circ f$ coincides with the set of optimal decision strategies for $b \circ g$ over g , where $g = a \circ f$:*

$$\text{Opt}_f(a \circ f) = \text{Opt}_{b \circ g}(g).$$

In a wide class of decision problems (e.g., in linear estimation problems) an optimal IT r happens to be deterministic and is specified by the “deterministic part” of the IT b .

For many categories of information transformers (for example, stochastic, multivalued, and fuzzy ITs [20, 12, 14]) an optimal decision strategy r can be constructed “pointwise” according to the following scheme. For the given “result of observation” $y \in \mathcal{R}$ consider the conditional (posterior) distribution $b(y)$ for f under a fixed $g = y$, and put

$$r(y) \stackrel{\text{def}}{=} d_{b(y)},$$

where $d_{b(y)}$ is an optimal decision with respect to the posterior distribution $b(y)$.

8. EXAMPLES OF CATEGORIES OF INFORMATION TRANSFORMERS

In this section we present several examples of different classes of information transformers. The major difference between them is the way of representing uncertainty. In each case (except the category stochastic linear ITs, which cannot be constructed as a Kleisli category, but is a subcategory of one) we will mention the corresponding: base category \mathbf{D} , functor T , and natural transformation γ . “Elementary” definitions for these categories may be found in [17, 13].

8.1. Stochastic ITs

Let $\mathbf{D} = \mathbf{Meas}$, the category of measurable spaces and measurable maps, $T\mathcal{A}$ is the space of all probability measures on \mathcal{A} , (details may be found in [25]) and $\gamma_{\mathcal{A},\mathcal{B}}$ takes a pair of distributions \mathcal{P}, \mathcal{Q} to their product $\mathcal{P} \otimes \mathcal{Q}$, a distribution on $\mathcal{A} \times \mathcal{B}$.

The category of stochastic information transformers \mathbf{ST} consists of measurable spaces (as objects) and transition probability functions (as morphisms, that is, information transformers) [3, 4, 25]. Note that a classical statistical experiment (namely, a parametrized family of probability measures), a statistics (namely, a measurable function of a sample of observations), and a decision strategy (possibly, nondeterministic) can be represented by appropriate transition probability functions. Thus, all the above concepts fit in well with this scheme.

Suppose $\mathcal{A} = \langle \Omega_{\mathcal{A}}, \mathfrak{S}_{\mathcal{A}} \rangle$ and $\mathcal{B} = \langle \Omega_{\mathcal{B}}, \mathfrak{S}_{\mathcal{B}} \rangle$ are two measurable spaces. A *stochastic information transformer* $a: \mathcal{A} \rightarrow \mathcal{B}$ is determined by a real-valued function (transition probability function [3, 26, 27]) $P_a(\omega, B)$ of two arguments $\omega \in \Omega_{\mathcal{A}}, B \in \mathfrak{S}_{\mathcal{B}}$ that satisfy the following conditions:

- (a) Given a fixed event $B \in \mathfrak{S}_{\mathcal{B}}$, the map $P_a(\cdot, B)$ is a measurable function on $\Omega_{\mathcal{A}}$.
- (b) Given a fixed elementary event $\omega \in \Omega_{\mathcal{A}}$, the map $P_a(\omega, \cdot)$ is a probability measure on $\langle \Omega_{\mathcal{B}}, \mathfrak{S}_{\mathcal{B}} \rangle$.

For a given stochastic information transformers $a: \mathcal{A} \rightarrow \mathcal{B}$ and $b: \mathcal{B} \rightarrow \mathcal{C}$ their *composition* $b \circ a$ in the category \mathbf{ST} corresponds to the transition probability function (see [3, 25])

$$P_{b \circ a}(\omega, C) \stackrel{\text{def}}{=} \int_{\Omega_{\mathcal{B}}} P_b(\omega', C) P_a(\omega, d\omega') \quad \forall \omega \in \Omega_{\mathcal{A}}, \quad \forall C \in \mathfrak{S}_{\mathcal{C}}.$$

The subcategory of *deterministic* ITs is actually a category \mathbf{Meas} of measurable spaces and measurable maps. To every measurable map $\varphi: \mathcal{A} \rightarrow \mathcal{B}$ there corresponds the transition probability function

$$P_{\varphi}(\omega, B) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } \varphi(\omega) \in B, \\ 0, & \text{if } \varphi(\omega) \notin B \end{cases} \quad \forall \omega \in \Omega_{\mathcal{A}}, \quad \forall B \in \mathfrak{S}_{\mathcal{B}}.$$

The category \mathbf{Meas} has products, namely, the product of measurable spaces \mathcal{A} and \mathcal{B} in \mathbf{Meas} is $\langle \mathcal{A} \times \mathcal{B}, \pi_{\mathcal{A},\mathcal{B}}, \nu_{\mathcal{A},\mathcal{B}} \rangle$, where

$$\mathcal{A} \times \mathcal{B} \stackrel{\text{def}}{=} \langle \Omega_{\mathcal{A}} \times \Omega_{\mathcal{B}}, \mathfrak{S}_{\mathcal{A}} \otimes \mathfrak{S}_{\mathcal{B}} \rangle,$$

$\pi_{\mathcal{A},\mathcal{B}}$ and $\nu_{\mathcal{A},\mathcal{B}}$ are the projections from the Cartesian product $\Omega_{\mathcal{A}} \times \Omega_{\mathcal{B}}$ onto its components $\Omega_{\mathcal{A}}$ and $\Omega_{\mathcal{B}}$ respectively, and $\mathfrak{S}_{\mathcal{A}} \otimes \mathfrak{S}_{\mathcal{B}}$ is the product of σ -algebras $\mathfrak{S}_{\mathcal{A}}$ and $\mathfrak{S}_{\mathcal{B}}$.

For a given pair of ITs $a: \mathcal{D} \rightarrow \mathcal{A}$ and $b: \mathcal{D} \rightarrow \mathcal{B}$ with a common source we define their *product* $a * b: \mathcal{D} \rightarrow \mathcal{A} \times \mathcal{B}$ so that for every $\omega \in \Omega_{\mathcal{D}}$ the probability distribution $P_{a*b}(\omega, \cdot)$ on $\mathcal{A} \times \mathcal{B}$ is the product \otimes of the distributions $P_a(\omega, \cdot)$ and $P_b(\omega, \cdot)$, that is,

$$P_{a*b}(\omega, \cdot) \stackrel{\text{def}}{=} P_a(\omega, \cdot) \otimes P_b(\omega, \cdot) \quad \forall \omega \in \Omega_{\mathcal{D}}.$$

In other words (see, for example, [26]), the distribution P_{a*b} is completely determined by the following condition:

$$P_{a*b}(\omega, A \times B) \stackrel{\text{def}}{=} P_a(\omega, A) P_b(\omega, B) \quad \forall \omega \in \Omega_{\mathcal{D}}, \quad \forall A \in \mathfrak{S}_{\mathcal{A}}, \quad \forall B \in \mathfrak{S}_{\mathcal{B}}.$$

The only obvious choice for the accuracy relation in the category of stochastic ITs seems to be the equality relation.

Now let us demonstrate that the basic concepts of mathematical statistics are adequately described in terms of this IT-category. Namely, we shall verify that the concepts of distribution, conditional distribution,

etc. (introduced above in terms of IT-categories), in the category of stochastic ITs lead to the corresponding classical concepts.

Indeed, any probability distribution Q on a given measurable space $\mathcal{A} = \langle \Omega_{\mathcal{A}}, \mathfrak{S}_{\mathcal{A}} \rangle$ is uniquely determined by the morphism $f: \mathcal{Z} \rightarrow \mathcal{A}$ from the terminal object $\mathcal{Z} = \langle \{0\}, \{\emptyset, \{0\}\} \rangle$ (a one-point measurable space) such that

$$P_f(0, A) = Q(A) \quad \forall A \in \mathfrak{S}_{\mathcal{A}}.$$

In what follows we shall omit the first argument in $P_f(0, A)$ and write just $P_f(A)$ instead.

A statistical experiment is described by a family of probability measures Q_{θ} on some measurable space \mathcal{B} . This family is usually parametrized by elements of a certain set $\Omega_{\mathcal{A}}$. Sometimes (especially when statistical problems with a prior information are studied) it is additionally assumed that the set $\Omega_{\mathcal{A}}$ is equipped by some σ -algebra $\mathfrak{S}_{\mathcal{A}}$ and that $Q_{\theta}(B)$ is a measurable function of $\theta \in \Omega_{\mathcal{A}}$ for all $B \in \mathfrak{S}_{\mathcal{B}}$ (and thus, $Q_{\theta}(B)$ is a transition probability function [27]). Therefore, such statistical experiment is determined by the stochastic information transformer $a: \mathcal{A} \rightarrow \mathcal{B}$, where

$$P_a(\theta, B) = Q_{\theta}(B) \quad \forall \theta \in \Omega_{\mathcal{A}}, \quad \forall B \in \mathfrak{S}_{\mathcal{B}}.$$

In the case when no σ -algebra on the set $\Omega_{\mathcal{A}}$ is specified, one can put $\mathfrak{S}_{\mathcal{A}} = \mathcal{P}(\Omega_{\mathcal{A}})$, that is, the σ -algebra of all the subsets of the set $\Omega_{\mathcal{A}}$. It is clear that in this case the function $P_a(\theta, B) = Q_{\theta}(B)$ is a measurable function of $\theta \in \Omega_{\mathcal{A}}$ for every fixed $B \in \mathfrak{S}_{\mathcal{B}}$ and thus (being a transition probability function), is described by a stochastic IT $a: \mathcal{A} \rightarrow \mathcal{B}$.

Note also, that any statistic, being a measurable function, is represented by a certain deterministic IT. Decision strategies also correspond to deterministic ITs. At the same time, nondeterministic (mixed) decision strategies are adequately represented by stochastic information transformers of general kind.

Now, let f be some fixed distribution on \mathcal{A} and let $a: \mathcal{A} \rightarrow \mathcal{B}$ be some IT. The joint distribution h on $\mathcal{A} \times \mathcal{B}$, generated by f and a (from the IT-categorical point of view, see Section 7) is

$$h = (i * a) \circ f.$$

It means that for every set $A \times B$, where $A \in \Omega_{\mathcal{A}}$ and $B \in \Omega_{\mathcal{B}}$,

$$\begin{aligned} P_h(A \times B) &= \int_{\Omega_{\mathcal{A}}} P_{i*a}(\omega, A \times B) P_f(d\omega) \\ &= \int_{\Omega_{\mathcal{A}}} P_i(\omega, A) P_a(\omega, B) P_f(d\omega) \\ &= \int_A P_a(\omega, B) P_f(d\omega). \end{aligned}$$

Thus we come to the well known classical expression for the generated joint distribution (see, for example, [27]).

Now assume that P_f is considered as some probability *prior* distribution (or distribution *a priori*) on \mathcal{A} . Then for a given transition probability function P_a , a *posterior* (or *conditional*) distribution $P_b(\omega', \cdot)$ on \mathcal{A} for a fixed $\omega' \in \Omega_{\mathcal{B}}$ is determined, accordingly to [27] by a transition probability function $P_b(\omega', A)$, $\omega' \in \Omega_{\mathcal{B}}$, $A \in \mathfrak{S}_{\mathcal{A}}$ such that

$$P_h(A \times B) = \int_B P_b(\omega', A) P_g(d\omega') \quad \forall A \in \mathfrak{S}_{\mathcal{A}}, \quad \forall B \in \mathfrak{S}_{\mathcal{B}},$$

where

$$P_g(B) = \int_{\Omega, \mathcal{A}} P_a(\omega, B) P_f(d\omega) \quad \forall B \in \mathfrak{G}_{\mathcal{B}}.$$

It is easily verified that in terms of ITs the above expressions have the following forms:

$$h = (b * i) \circ g,$$

where

$$g = a \circ f.$$

This shows, that the classical concept of conditional distribution is adequately described by the concept of conditional IT in terms of categories of information transformers.

8.2. Linear stochastic ITs with additive noise

As we will see this category of ITs cannot be constructed as a Kleisli category, but is a subcategory the category of stochastic ITs, examined above.

Suppose \mathcal{D} and \mathcal{R} are arbitrary finite-dimensional Euclidean spaces. We shall say that a linear information transformer [10, 11] (measurement model [28]) a acting from \mathcal{D} to \mathcal{R}

$$a: \mathcal{D} \rightarrow \mathcal{R},$$

is determined by a pair

$$\langle A_a, \Sigma_a \rangle, \quad A_a: \mathcal{D} \rightarrow \mathcal{R}, \quad \Sigma_a: \mathcal{R} \rightarrow \mathcal{R}, \quad \Sigma_a \geq 0,$$

where A_a and Σ_a are linear maps.

Such pair $\langle A_a, \Sigma_a \rangle$ represents a statistical experiment of the form [28]

$$y = A_a x + \nu, \quad x \in \mathcal{D}, \quad y \in \mathcal{R},$$

where ν is a random vector in \mathcal{R} with the zero mean and the correlation operator Σ_a .

The *composition* of two linear ITs $\langle A_a, \Sigma_a \rangle: \mathcal{D} \rightarrow \mathcal{A}$ and $\langle A_b, \Sigma_b \rangle: \mathcal{A} \rightarrow \mathcal{B}$ is defined by

$$\langle A_b, \Sigma_b \rangle \circ \langle A_a, \Sigma_a \rangle \stackrel{\text{def}}{=} \langle A_b A_a, \Sigma_b + A_b \Sigma_a A_b^* \rangle.$$

The composition corresponds to the consecutive connection of information transformers that have independent random errors.

The *product* of two information transformers

$$\langle A_a, \Sigma_a \rangle: \mathcal{D} \rightarrow \mathcal{A}, \quad \langle A_b, \Sigma_b \rangle: \mathcal{D} \rightarrow \mathcal{B}$$

is defined by:

$$\langle A_a, \Sigma_a \rangle * \langle A_b, \Sigma_b \rangle \stackrel{\text{def}}{=} \langle A_{a*b}, \Sigma_{a*b} \rangle: \mathcal{D} \rightarrow \mathcal{A} \times \mathcal{B},$$

where

$$\begin{aligned} A_{a*b}: \mathcal{D} \rightarrow \mathcal{A} \times \mathcal{B}, \quad A_{a*b} x &\stackrel{\text{def}}{=} \langle A_a x, A_b x \rangle, \\ \Sigma_{a*b}: \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{A} \times \mathcal{B}, \quad \Sigma_{a*b} \langle x, y \rangle &\stackrel{\text{def}}{=} \langle \Sigma_a x, \Sigma_b y \rangle. \end{aligned}$$

This construction gives us the category **SLT** with the subcategory of deterministic ITs is (isomorphic to) the category of Euclidean spaces and linear maps. In this case a linear map $A: \mathcal{D} \rightarrow \mathcal{R}$ corresponds to the IT $\langle A, 0 \rangle: \mathcal{D} \rightarrow \mathcal{R}$.

As we have already mentioned this category of ITs cannot be constructed as a Kleisli category over the category of finite dimensional Euclidean spaces. Indeed we can not define a “space of distributions” on some space \mathcal{A} as a finite dimensional linear space, and thus, can not define functor T in the category of finite dimensional Euclidean spaces. However, **SLT** may be considered as a subcategory the category of stochastic ITs **ST**, examined above. Indeed, each Euclidean space may be considered as a measurable space endowed with Borel σ -algebra. Finally, we may consider an IT $a = \langle A_a, \Sigma_a \rangle : \mathcal{D} \rightarrow \mathcal{R}$ (in **SLT**) as the transition probability, that takes an element $x \in \mathcal{D}$ to the normal distribution $N(A_ax, \Sigma_a)$ with the mean value A_ax and the correlation operator Σ_a . Routine verification shows, that the composition and product operations are preserved under such inclusion of **SLT** into **ST**.

In addition to the trivial relation of accuracy (which coincides with the equality relation) one can define the accuracy relation in the following way:

$$\langle A_a, \Sigma_a \rangle \triangleright \langle A_b, \Sigma_b \rangle \stackrel{\text{def}}{\iff} A_a = A_b, \Sigma_a \leq \Sigma_b.$$

However, it can be proved that the informativeness relations corresponding these different accuracy preorders, actually coincide.

In the category of linear information transformers every equivalence class $[a]$ corresponds to a pair $\langle Q, S \rangle$, where $Q \subseteq \mathcal{D}$ is an Euclidean subspace and $S : Q \rightarrow Q$ is nonnegative definite operator, that is, $S \geq 0$. In these terms

$$\langle Q_1, S_1 \rangle \geq \langle Q_2, S_2 \rangle \stackrel{\text{def}}{\iff} Q_1 \supseteq Q_2, S_1 \upharpoonright Q_2 \leq S_2.$$

Here $S_1 \upharpoonright Q_2$ (the restriction of S_1 on Q_2) is defined by the expression $S_1 \upharpoonright Q_2 \stackrel{\text{def}}{=} P_2 I_1 S_1 P_1 I_2$, where $I_j : Q_j \rightarrow \mathcal{D}$ is the subspace inclusion, and $P_j : \mathcal{D} \rightarrow Q_j$ is the orthogonal projection (cf. [10, 28]).

Note also that in the category of linear information transformers every IT is dominated (in the sense of the preorder relation \triangleright) by a deterministic IT. Hence, according to Proposition 1, in any monotone decision-making problem without loss of quality one can search optimal decision strategies in the class of deterministic ITs.

It is shown in [11], that in the category of linear ITs for any joint distribution there always exist conditional distributions. Thus in problems with a prior information one can apply Bayesian principle. Its direct proof in the category of linear ITs as well as the explicit expression for conditional information transformers can be found in [11].

8.3. The category of sets as a category of ITs

As a trivial example of IT-category we consider the category of sets **Set**, whose objects are sets and morphisms are maps. This category has products, hence all the ITs are deterministic. In fact this category is trivially a Kleisli category with identity functor as functor T .

It is not hard to prove that for a given set \mathcal{D} , the class of equivalent informativeness for an IT a with the set \mathcal{D} being its domain, is completely determined by the following equivalence relation \approx_a on \mathcal{D} :

$$x \approx_a y \stackrel{\text{def}}{\iff} ax = ay \quad \forall x, y \in \mathcal{D}.$$

Furthermore, $a \succ b$ if and only if the equivalence relation \approx_a is *finer* than \approx_b , that is,

$$a \succ b \iff \forall x, y \in \mathcal{D} \quad (x \approx_a y \implies x \approx_b y).$$

Thus, the partially ordered monoid of equivalence classes for ITs with the source \mathcal{D} , is isomorphic to the monoid of all equivalence relations on \mathcal{D} equipped with the order “finer” and with the product:

$$x (\approx_a * \approx_b) y \stackrel{\text{def}}{\iff} (x \approx_a y, x \approx_b y) \quad \forall x, y \in \mathcal{D}.$$

8.4. Multivalued ITs

Let $\mathbf{D} = \mathbf{Set}$, the category of sets, $T\mathcal{A}$ is the set of all nonempty subsets of \mathcal{A} and $\gamma_{\mathcal{A},\mathcal{B}}$ takes pair of sets \mathcal{P}, \mathcal{Q} to their Cartesian product $\mathcal{P} \times \mathcal{Q}$, a subset of $\mathcal{A} \times \mathcal{B}$. This leads us to the category \mathbf{MVT} of multivalued ITs. Detailed study of this category may be found in [13]. Thus, the category \mathbf{MVT} consists of sets as objects and of multivalued maps (everywhere defined relations) as morphisms (information transformers). Despite its simplicity, this class of ITs may be convenient when stochastic description of measurement error is inadequate.

So, a multivalued IT a from \mathcal{D} to \mathcal{R}

$$a: \mathcal{D} \rightarrow \mathcal{R}$$

is determined by a multivalued map, that is,

$$\forall x \in \mathcal{D} \quad ax \subseteq \mathcal{R}, \quad ax \neq \emptyset.$$

Define the *composition* and the *product* of multivalued ITs by the following expressions:

$$(b \circ a)(x) \stackrel{\text{def}}{=} \bigcup \{by \mid y \in ax\},$$

$$(a * b)(x) \stackrel{\text{def}}{=} ax \times bx.$$

The subcategory of deterministic ITs is actually the category of sets \mathbf{Set} .

In addition to the trivial accuracy relation in the category of multivalued ITs one can put

$$a \triangleright b \stackrel{\text{def}}{\iff} \forall x \in \mathcal{D} \quad ax \subseteq bx.$$

These two accuracy relations lead to different informativeness relations [13], called (strong) informativeness \succcurlyeq and weak informativeness $\dot{\succcurlyeq}$.

For the both informativeness relations the classes of equivalent ITs with a fixed source \mathcal{D} can be described explicitly.

In the case of weak informativeness every class of equivalent ITs corresponds to a certain covering \mathcal{P} of the set \mathcal{D} , such that if \mathcal{P} contains some set B then it contains all its subsets:

$$(\exists B \in \mathcal{P} (A \subseteq B)) \implies A \in \mathcal{P}.$$

Moreover, a covering \mathcal{P}_1 is more (weakly) informative than \mathcal{P}_2 (namely, \mathcal{P}_1 corresponds to a class of more (weakly) informative ITs than \mathcal{P}_2) if \mathcal{P}_1 is contained in \mathcal{P}_2 , that is,

$$\mathcal{P}_1 \dot{\succcurlyeq} \mathcal{P}_2 \stackrel{\text{def}}{\iff} \mathcal{P}_1 \subseteq \mathcal{P}_2.$$

In the case of (strong) informativeness every class of equivalent ITs corresponds to a covering \mathcal{P} of the set \mathcal{D} , that satisfy the more complex condition:

$$\left((\exists B \in \mathcal{P} A \subseteq B) \ \& \ (\exists \mathcal{B} \subseteq \mathcal{P} A = \bigcup \mathcal{B}) \right) \implies A \in \mathcal{P}.$$

In this case

$$\mathcal{P}_1 \succcurlyeq \mathcal{P}_2 \stackrel{\text{def}}{\iff} \left((\forall A \in \mathcal{P}_1 \ \exists B \in \mathcal{P}_2 \ A \subseteq B) \ \& \ (\forall B \in \mathcal{P}_2 \ \exists \mathcal{A} \subseteq \mathcal{P}_1 \ B = \bigcup \mathcal{A}) \right).$$

In the category of multivalued information transformers every IT is dominated (in the sense of the partial order \triangleright) by a deterministic IT. Thus, in the monotone decision-making problem one can search optimal decision strategies in the class of deterministic ones.

For every joint distribution in the category of multivalued ITs there exist conditional distributions [12]. Therefore, in decision problems with a prior information, the Bayesian approach can be effectively applied.

8.5. Categories of fuzzy information transformers

Here we define two categories of fuzzy information transformers **FMT** and **FPT** that correspond to different fuzzy theories [14]. Let $\mathbf{D} = \mathbf{Set}$, $T\mathcal{A}$ is the set of all normalized fuzzy subsets of \mathcal{A} and $\gamma_{\mathcal{A},\mathcal{B}}$ takes pair of fuzzy subsets \mathcal{P}, \mathcal{Q} to the fuzzy subset $\mathcal{P} \times \mathcal{Q}$, $(\mathcal{P} \times \mathcal{Q})(x, y) = \mathcal{P}(x) \otimes \mathcal{Q}(y)$, where the operation \otimes may be defined in a variety of ways. The most common are the minimum (the category **FMT**) and product (the category **FPT**) operations [14].

Objects of these categories are arbitrary sets and morphisms are everywhere defined fuzzy maps, namely, maps that take an element to a normed fuzzy set (a fuzzy set A is normed if supremum of its membership function μ_A is 1). Thus, an information transformer $a: \mathcal{A} \rightarrow \mathcal{B}$ is defined by a membership function $\mu_{ax}(y)$ which is interpreted as the grade of membership of an element $y \in \mathcal{B}$ to a fuzzy set ax for every element $x \in \mathcal{A}$.

The category **FMT**. Suppose $a: \mathcal{A} \rightarrow \mathcal{B}$ and $b: \mathcal{B} \rightarrow \mathcal{C}$ are some fuzzy maps. We define their *composition* $b \circ a$ as follows: for every element $x \in \mathcal{A}$ put

$$\mu_{(b \circ a)x}(z) \stackrel{\text{def}}{=} \sup_{y \in \mathcal{B}} \min \left(\mu_{ax}(y), \mu_{by}(z) \right).$$

For a pairs of fuzzy information transformers $a: \mathcal{D} \rightarrow \mathcal{A}$ and $b: \mathcal{D} \rightarrow \mathcal{B}$ with the common source \mathcal{D} , we define their *product* as the IT that acts from \mathcal{D} to the Cartesian product $\mathcal{A} \times \mathcal{B}$, such that

$$\mu_{(a * b)x}(y, z) \stackrel{\text{def}}{=} \min \left(\mu_{ax}(y), \mu_{bx}(z) \right).$$

The category **FPT**. Define the *composition* and the *product* by the following expressions:

$$\mu_{(b \circ a)x}(z) \stackrel{\text{def}}{=} \sup_{y \in \mathcal{B}} \left(\mu_{ax}(y) \mu_{by}(z) \right),$$

$$\mu_{(a * b)x}(y, z) \stackrel{\text{def}}{=} \mu_{ax}(y) \mu_{bx}(z).$$

In the both defined above categories of fuzzy information transformers the subcategory of *deterministic* ITs is (isomorphic to) the category of sets **Set**. Let $g: \mathcal{A} \rightarrow \mathcal{B}$ be some map (morphism in **Set**). Define the corresponding fuzzy IT (namely, a fuzzy map, which is obviously, everywhere defined) $\tilde{g}: \mathcal{A} \rightarrow \mathcal{B}$ in the following way:

$$\mu_{\tilde{g}(x)}(y) \stackrel{\text{def}}{=} \delta_{g(x),y} = \begin{cases} 1, & \text{if } g(x) = y, \\ 0, & \text{if } g(x) \neq y. \end{cases}$$

Concerning the choice of accuracy relation, note, that in these IT-categories, like in the category of multivalued ITs, apart from the trivial accuracy relation one can put for $a, b: \mathcal{A} \rightarrow \mathcal{B}$

$$a \triangleright b \stackrel{\text{def}}{\iff} \forall x \in \mathcal{A} \quad \forall y \in \mathcal{B} \quad \mu_{ax}(y) \leq \mu_{bx}(y).$$

In each fuzzy IT-category these two choices lead to two different informativeness relations, namely the strong and the weak ones.

Like in the categories of linear and multivalued ITs discussed above, monotone decision-making problems admit restriction of the class of optimal decision strategies to deterministic ITs without loss of quality.

It was shown in [14] that for every joint distribution in the categories of fuzzy ITs there exist conditional distributions. It allows Bayesian approach and makes use of Bayesian principle in decision problems with a prior information for fuzzy ITs [14] (see also [15], where connections between fuzzy decision problems and the underlying fuzzy logic are studied).

In this section we introduced only several examples of IT-categories. Let us also remark that there is an extensive literature that studies a wide spectrum of categories which are close in their structure to IT-categories [29, 30, 31, 32, 33, 34, 23].

REFERENCES

1. Sacksteder R. A Note on Statistical Equivalence. *Ann. Math. Statist.*, 1967, vol. 38, no. 3, pp. 787–795.
2. Morse N., Sacksteder R. Statistical Isomorphism. *Ann. Math. Statist.*, 1966, vol. 37, no. 2, pp. 203–214.
3. Chentsov N.N. *Statistical Decision Rules and Optimal Inference* [in Russian]. Moscow: Nauka, 1972.
4. Chentsov N.N. Categories of Mathematical Statistics. *Dokl. Akad. Nauk SSSR*, 1965, vol. 164, no. 3, pp. 511–514.
5. Golubtsov P.V. Measurement Systems: Algebraic Properties and Informativity. *Pattern Recognition and Image Analysis*, 1991, vol. 1, no. 1, pp. 77–86.
6. MacLane S. *Categories for the Working Mathematician*. New York: Springer, 1971.
7. Herrlich H., Strecker G.E. *Category Theory*. Boston: Allyn and Bacon, 1973.
8. Arbib M.A., Manes E.G. *Arrows, Structures and Functors*. New York: Academic Press, 1975.
9. Goldblatt R. *Topoi. The Categorical Analysis of Logic*. Amsterdam: North-Holland, 1979.
10. Golubtsov P.V. Informativity in the Category of Linear Measurement Systems. *Probl. Inf. Transm.*, 1992, vol. 28, no. 2, pp. 125–140.
11. Golubtsov P.V. Relative Informativity and a Priori Information in the Category of Linear Information Transformers. *Probl. Inf. Transm.*, 1995, vol. 31, no. 3, pp. 195–215.
12. Golubtsov P.V., Filatova S.A. Multivalued Measurement-Computer Systems. *Mat. Model.*, 1992, vol. 4, no. 7, pp. 79–94.
13. Golubtsov P.V. Informativity in the Category of Multivalued Information Transformers. *Probl. Inf. Transm.*, 1998, vol. 34, no. 3, pp. 259–276.
14. Golubtsov P.V. Theory of Fuzzy Sets as a Theory of Uncertainty and Decision-Making Problems in Fuzzy Experiments. *Probl. Inf. Transm.*, 1994, vol. 30, no. 3, pp. 232–250.
15. Golubtsov P.V. Fuzzy Logical Semantics of Bayesian Decision Making. *Applications of Fuzzy Logic Technology II. SPIE Proceedings. Vol. 2493.*, Orlando, Florida, 1995, pp. 228–239.
16. Golubtsov P.V. Categories of Information Transformers and the Concept of Informativity. *Proc. Int. Conf. on Informatics and Control (ICI&C'97). Vol. 2*, St.-Petersburg, Russia, 1997., pp. 512–517.
17. Golubtsov P.V. Axiomatic Description of Categories of Information Transformers. *Probl. Inf. Transm.*, 1999, vol. 35, no. 3, pp. 80–99.
18. Kleisli H. Every Standard Construction is Induced by a Pair of Adjoint Functors. *Proc. Amer. Math. Soc.*, 1965, vol. 16, pp. 544–546.
19. Barr M., Wells C. *Toposes, Triples and Theories*. New York: Springer, 1984.
20. Borovkov A.A. *Mathematical Statistics. Supplementary Chapters*. [in Russian]. Moscow: Nauka, 1984.
21. Blackwell D. Comparison of Experiments. *Proc. Second Berkeley Sympos. on Mathematical Statistics and Probability.*, University of California Press, 1951, pp. 93–102.
22. Blackwell D. Equivalent Comparison of Experiments. *Ann. Math. Statist.*, 1953, vol. 24, no. 2, pp. 265–272.
23. Manes E.G. A Class of Fuzzy Theories. *J. Math. Anal. and Appl.*, 1982, vol. 85, pp. 409–451.
24. Blackwell D., Girschick M.A. *Theory of Games and Statistical Decisions*. New York: Wiley Sons, 1954.
25. Giry M. A Categorical Approach to Probability Theory. *Categorical Aspects of Topology and Analysis. Lecture Notes in Mathematics. No. 915.*, Berlin: Springer-Verlag, 1982, pp. 68–85.
26. Neveu J. *Bases Mathematiques du Calculus des Probabilites*. Paris: Masson et Cie, 1964.
27. Barra J.-R. *Notions Fondamentales de Statistique Mathematique*. Paris: Dunod, 1974.
28. Pyt'ev Yu.P. Reduction Problems in Experimental Research. *Mat. Sb.*, 1983, vol. 120, no. 2, pp. 240–272.
29. Goguen J.A. L-Fuzzy Sets. *J. Math. Anal. and Appl.*, 1967, vol. 18, pp. 145–174.

30. Goguen J.A. Categories of V-Sets. *Bull. Amer. Math. Soc.*, 1969, vol. 75, pp. 622–624.
31. Goguen J.A. Concept Representation in Natural and Artificial Languages: Axioms, Extensions, and Applications for Fuzzy Sets. *Int. J. Man-Machine Studies*, 1974, vol. 6, pp. 513–561.
32. Arbib M.A., Manes E.G. Fuzzy Machines in a Category. *Bull. Austral. Math. Soc.*, 1975, vol. 13, pp. 169–210.
33. Arbib M.A., Manes E.G. Fuzzy Morphisms in Automata Theory. *Category Theory Applied to Computation and Control. Lecture Notes in Computer Science. No. 25.*, New York: Springer-Verlag, 1976, pp. 80–86.
34. Manes E.G. *Algebraic Theories*. New York: Springer, 1976.

This paper was recommended for publication by V.A.Lyubetsky, a member of the Editorial Board