An introduction to the category of chance

Claudio Sossai
ISIB-CNR, Corso Stati Uniti 4, I-35127 Padova, Italy
claudio.sossai@isib.cnr.it

Abstract

Using category theory a mathematical analysis of chance is presented. Laws of chance are defined as properties that are time and uncertainty invariants. It is shown that proofs, interpreted into sets of observables, are laws of chance.

Key words: proof theory, topos theory, reasoning under uncertainty.

1 Introduction

Although the motivations of this work are foundational, we believe that there are many links with arguments of interest in the field of Information Processing and Management of Uncertainty. [1] This is due to the following observation: chance is at the root of information processing under uncertainty, and the analysis of the symmetries of chance is the cornerstone of the work. Uncertainty can be seen as the theory of limit properties of chance, for instance a probability measure is associated to an infinite set of data generated by chance. In many cases some explicit knowledge about the probability measure that governs the process is assumed.

Aim of this work is to show that the rules of logic are laws of chance. To prove this we must give a mathematical definition of law of chance.

[1] This paper originates from the suggestions of the reviewers of three papers [6, 7, 8] submitted to the IPMU-08 conference. Aim of this work is to give, in an informal way, the motivations and explanations of the theory presented in the three papers, where the interested reader can find all the mathematical definitions and proofs. [6, 7, 8] can be downloaded at the following address:http://www.isib.cnr.it/infor/papers/stat.pdf

Once we arrive at this definition, we discover a method that computes directly with finite sets of data generated by chance without any assumption about the unknown measure μ_0 that governs the process. Such method, as we will see, is strong enough to prove its own consistency.

The following facts are assumed:

1) to every finite set of data x we can associate an empirical measure σ_x that does not necessarily coincide with the measure μ_0.

2) at the limit, i.e. if we could know an infinite set of data, the measures associated to the infinite sets of data coincide with μ_0. Nonetheless infinite sets of data are not available, hence μ_0 remains always unknown.

For instance assume that x is generated by a coin-tossing experiment, then the Glivenko-Cantelli theorem gives us the method that satisfies the above facts. Nonetheless if we cannot assume that the coin is fair, using a finite set of outcomes we do not know μ_0. Moreover, in many cases, the empirical measure associated to x does not coincide with μ_0. Even worse, using only a finite number of outcomes, and without any hypothesis about μ_0, we do not know whether at least one of the observed empirical measures coincides with μ_0. This is the typical situation analyzed in this work: how can we compute, in a sound way, with the available data, i.e. finite sets of outcomes, assuming nothing about the measure μ_0?

2 Proofs and observables

To find a method to compute in a sound way with finite sets of data (without any knowledge about

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 204–211
Torremolinos (Málaga), June 22–27, 2008

μ_0) we must start from some general properties of chance that hold for every set of data x and whatever will be the measure μ_0.

A first general property of chance stems from the following observation: information coming from finite sets of observables has a *polarity*. Whatever will be μ_0, for every x there are only two possibilities:

1) x gives positive information w.r.t. μ_0 if σ_x approximates μ_0, i.e. $\mu_0 = \sigma_x$

2) x gives negative information w.r.t. μ_0 if σ_x does not approximate μ_0, i.e. $\mu_0 \neq \sigma_x$.

Given a pair of observables x, y the same situation applies: x gives positive information w.r.t y if $\sigma_x = \sigma_y$ (shortly $xSTATy$), otherwise x gives negative information w.r.t y if $\sigma_x \neq \sigma_y$ (shortly $xTESTy$).

From this simple observation we can derive the fundamental relation: two observables x, y, are positively-coherent iff $xSTATy$ otherwise they are negatively-coherent iff $xTESTy$.

This is not a novelty, in fact in statistics we have methods that use the available data to approximate a measure and other methods, usually called tests, that use the available data in a dual way, i.e. to exclude a measure.

The novelty, to our knowledge, stays in a method that combines the two sources of information: positive and negative.

To find this method we need logic. More precisely linear logic, because, as we will see, in linear logic the link between reasoning (proof theory) and computing (typed lambda calculus) is at the root of the formalism. This link is the key for a method that computes directly with data. It is worth noting that classical and intuitionistic logic can be translated into linear logic (see [3, 4]).

To proceed in the analysis we must specify what we mean by sets of data. To this aim let us introduce the concept of experiment.

An *experiment* is a stochastic process $X = \{X_n\}_{n \in N}$ with a method σ^X that associates to every trajectory x, or observable of the process, a measure σ^X_x. An observable x is a finite set of outcomes, i.e. $x = \{X_0(\xi), ...X_n(\xi)\}$. By set of data we mean an observable of a stochastic process.

To simplify the discussion we assume that all the random variables are measurable w.r.t. a Boolean field of sets $\mathcal{B}$ defined over a set Ξ.

To every stochastic process X it is associated a family of fields of sets over a set Ξ, $\mathcal{F}^X = \{\mathcal{F}^X_n\}_{n \in N}$, with $\mathcal{F}^X_n \leq \mathcal{F}^X_m$ if $n \leq m$, called filtration. It is a well known fact that the filtration $\mathcal{F}^X$ contains all the information of the process X (see [1]). The filtration can be a complex structure, but for simple processes we can give some examples. For instance, in figure 1 the filtration associated to a fair coin experiment is represented. To construct the filtration take $\Xi = [0, 1]$ and divide the interval into four equal parts $p_1,..p_4$. For $\xi \in p_1$ let $X_0(\xi) = h$, while for $\xi \in p_2$ let $X_0(\xi) = t$, and so on. This defines $\mathcal{F}^X_0$. For $\mathcal{F}^X_1$ divide every p_i into two equal parts and proceed as above.

Given two stochastic processes X, Y we can define a new process $X \times Y$, called the product process, described by the filtration obtained by the intersection of the two filtrations. It is not difficult to see that every $\mathcal{F}^X_n$ is atomic, the atoms being the sets $X^{-1}_n(X_n(\xi))$, $\xi \in \Xi$, hence we can define $\mathcal{F}^{X \times Y}_n$ as the algebra generated by the atoms $X^{-1}_n(X_n(\xi)) \cap Y^{-1}_n(Y_n(\xi))$, $\xi \in \Xi$. In figure 2 we have depicted the filtration obtained as the product experiment of a coin experiment with a urn experiment where the urn contains an equal number of red, white and black balls.

We can combine dependent or independent experiments. In this paper, due to the lack of space, we will analyze the case of independent experiments. The full case with dependent experiments can be found in [6, 7, 8].

If X and Y are independent then we can assume that the measure that governs the combined experiment is the product measure of the two measures that govern the single experiments. Therefore we obtain the definition of product experiment: the process is the product process and the method associates to every pair of observables xy the product measure, i.e. $\sigma^{X \times Y}_{xy} = \sigma^X_x \times \sigma^Y_y$.

Summing up we have assumed a *second general property of chance*: independent experiments are governed by the product measure.

Now we have a sufficient number of chance symmetries to discover a sound method that computes sets of observables from finite sets of observables.

For every experiment X the set of all observables will be denoted by $|X|$, moreover we assume that for every experiment we have observed $\overline{x} \in |X|$.

The key idea is the definition of *coherence* between observables. Here the definition is given in an intuitive form; for the exact definition see [6, 7, 8].

If P is an atomic experiment then $x, y \in |P|$ are coherent (written $xSTAT_P y$) iff $\sigma_x^P = \sigma_y^P$, while they are incoherent (written $xTEST_P y$) iff $\sigma_x^P \neq \sigma_y^P$. Negation is defined using the $TEST$ relation: for instance, the negation of P, written $P^\perp$, is defined as $xSTAT_{P^\perp} y$ iff $xTEST_P y$.

This is a novelty: we have a method that uses, in a unique framework, positive and negative information. It is also evident that logic is the natural environment where this important symmetry of chance can be represented.

For compound experiments let us recall that two product measures are equal iff their components are equal (sse [5]), hence $xySTAT_{X\times Y}x'y'$ iff $\sigma_{xy}^{X\times Y} = \sigma_{x'y'}^{X\times Y}$ iff $\sigma_x^X = \sigma_{x'}^X$ **and** $\sigma_y^Y = \sigma_{y'}^Y$ iff $xSTAT_X x'$ **and** $ySTAT_Y y'$.

Moreover $xyTEST_{X\times Y}x'y'$ iff $\sigma_{xy}^{X\times Y} \neq \sigma_{x'y'}^{X\times Y}$ iff $\sigma_x^X \neq \sigma_{x'}^X$ **or** $\sigma_y^Y \neq \sigma_{y'}^Y$ iff $xTEST_X x'$ **or** $yTEST_Y y'$.

This gives us a sufficient machinery to discover a logic behind the notion of coherence defined using the two above-mentioned general properties of chance.

The idea is the following: to every atomic proposition P we can associate an experiment $\epsilon(P)$ with the $STAT_{\epsilon(P)}$ relation; to the negation of the atomic proposition P we can associate the same experiment with the $TEST_{\epsilon(P)}$ relation, hence $xSTAT_{\epsilon(P^\perp)}x'$ iff $xTEST_{\epsilon(P)}x'$. To the formula $\phi \otimes \psi$ (i.e. ϕ **and** ψ) we can associate the product experiment with the corresponding coherent relation. To the formula $\phi ⅋ \psi$ (i.e. ϕ **or** ψ) we can associate the product experiment with the coherent relation defined as $xySTAT_{\epsilon(\phi ⅋ \psi)}x'y'$ iff $xySTAT_{\epsilon((\phi^\perp \otimes \psi^\perp)^\perp)}x'y'$. Implication is defined as: $\phi \multimap \psi = \phi^\perp ⅋ \psi$.

Using the above ideas we can associate to every formula ϕ an experiment $\epsilon(\phi)$ and a coherence relation $STAT_{\epsilon(\phi)}$. A set of observables of $\epsilon(\phi)$ that are coherent w.r.t. the available observable $\overline{x} \in \epsilon(\phi)$ is called a stochastic clique (denoted by $a_{\overline{x}}^{\phi}$). It is possible to prove (see [6, 7, 8]) that every proof π of a formula ϕ ($\vdash_\pi \phi$) constructs a stochastic clique. This shows that the objects that describe the above stochastic properties are the same as those that describe proofs in the denotational semantics of logic.

Now let us see how to construct a method, based on the above symmetries of chance, that transforms directly sets of observables into sets of observables in a sound way.

Assume that $\vdash_\pi \phi \multimap \psi$, then π constructs a stochastic clique $a_{\overline{xy}}^{\phi^\perp ⅋ \psi}$ made of observables that are coherent w.r.t. the available observables $\overline{x}$ and $\overline{y}$. Let b be a coherent set of observables of $|\epsilon(\phi)|$, i.e. for every x, x' in b it holds $xSTAT_{\epsilon(\phi)}x'$.

Let $c = \{y : y \in |\epsilon(\psi)| \wedge (\exists x)(xy \in a_{\overline{xy}}^{\phi^\perp ⅋ \psi})\}$. It is easy to see that c is a coherent set of observables of $|\epsilon(\psi)|$. In fact if $y, y' \in c$ then there exist x, x' s.t. $xy \in a_{\overline{xy}}^{\phi^\perp ⅋ \psi}$ and $x'y' \in a_{\overline{xy}}^{\phi^\perp ⅋ \psi}$. From this and from the fact that $a_{\overline{xy}}^{\phi^\perp ⅋ \psi}$ is a stochastic clique we have that $xySTAT_{\epsilon(\phi^\perp ⅋ \psi)}x'y'$ iff $xSTAT_{\epsilon(\phi^\perp)}x'$ or $ySTAT_{\epsilon(\psi)}y'$. But $x, x' \in b$ hence $xSTAT_{\epsilon(\phi)}x'$ i.e. not $xSTAT_{\epsilon(\phi^\perp)}x'$, and from this it follows that $ySTAT_{\epsilon(\phi)}y'$. Using this observation, we can construct a function f, defined, as above, $f(b) = c$, for every coherent set of observables b of $|\epsilon(\phi)|$. Such function f transforms coherent sets of observables into coherent sets of observables. These functions, called linear functions, give the name to linear logic. They are the bridge between logic and computation, in fact linear functions allow us to use implications as models of lambda terms of the form λxt.

3 The category of chance

We have seen that the above machinery is based on the available observables $\overline{x}, \overline{y}$.... Hence also negation and the connectives depend on the available information, for instance changing $\overline{x}$ the corresponding $xSTAT\overline{x}$ and $xTEST\overline{x}$ relations change and hence also the stochastic cliques $a^{\phi}_{\overline{x}}$ used to give meaning to proofs and from this all the machinery. We have assumed that the available information is given by finite sets of data, hence we have no idea of the measure that governs the process. For instance if we flip twice a coin and get ht, in this information state, we believe that the coin is fair and hence the probability of head equals the probability of tail. If we now flip again the coin, whhichever will be the outcome we will change our idea of the measure that governs the process.

At this point we have to prove that the method remains sound also in the presence of this variability of information. To show this we need a formal definition of law of chance, i.e. properties that do not depend on $\overline{x}$ or μ_0. For this definition we need category and topos theory. Let us try to justify the choice of this complex mathematical theory.

To show the soundness of the method it is sufficient to prove that the objects constructed by proofs, i.e. the interpretations of proofs, do not depend on the available information: the observed data. This can be done showing that these objects do not change if the available information varies, i.e. they are laws of chance. Therefore the starting point, from a mathematical perspective, is the description of chance as it acts or transforms the observables. Once we have this mathematical description it remains to show that the interpretations of proofs do not vary w.r.t. these transformations.

In category theory, mathematical properties are described up to morphisms, that means w.r.t. their transformations. For instance in algebra mathematical properties of groups are described defining the properties of group operations, i.e. associativity with neuter element... In the categorical definition of group all is defined w.r.t. the transformations that preserve the properties, i.e. morphisms. Therefore category theory is a good mathematical theory if we want to describe chance as it transforms the observables.

Let us start the analysis of the dynamic of observables from the duality between positive and negative information. To help the intuition let us fix an experiment X.

As we have assumed, every infinite number of outcomes $x_\infty \in |X|$ determines with certainty a measure. Therefore if we look from an infinite future to the present, every x_∞ determines a unique measure μ_0 that can govern the process. Every x_∞ divides all the possible observables into two classes, the set of statistics (i.e. the observables that generate measures in accordance with μ_0) and the set of tests (i.e. observables that generate the measures that are different from μ_0). Moreover a x_∞ completely determines a realization of the experiment X.

The knowledge of one x_∞ will be called the *condition of total information*. From this observation we can say that the relevant information is contained in the observables made by an infinite number of outcomes. These facts will be mathematically described by the category $\mathcal{T}^X$ of stochastic time of the experiment X. To help the intuition first let us give a rough description of $\mathcal{T}^X$.

As a first approximation the objects of $\mathcal{T}^X$ are the observables and there is an arrow $f : y \to x$ between two observables iff all the information of x is contained in y, i.e. x and y give the same statistical measure, and $x \subseteq y$.

Note that, in usual time, the condition $x \subseteq y$ is sufficient to say that y is in the future of x, while in $\mathcal{T}^X$, for the existence of a time arrow between y and x we must also require that the two measures associated to x and y are the same, i.e. in stochastic time, time arrows respect statistical coherence.

Due to the above observations, the arrows of the category $\mathcal{T}^X$ of stochastic time go from the future to the past preserving the measure observed in the future.

Note that in this description of time there can be no time arrow between ht and h as in linear time, because the shift from h to ht requires a change in information state. Under the hypothesis of the *soundness of the statistics*, the future is described by a sequence of coherent outcomes, i.e. they generate the same statistical measure. Therefore the future of ht is described by all observables x with $ht \subseteq x$ that satisfy the constraint: $\sigma_x(h) = \sigma_x(t) = \frac{1}{2}$, while the future of h is made by the observables y s.t. $\sigma_y(h) = 1$.

From a stochastic point of view the two futures refer to two different realizations of the experiment X: in the first case the realization is generated by an unbiased coin, while in the second case we have a completely biased coin.

For this reason, in the description of the evolution of an experiment, incoherent time evolutions are not allowed, because here time is linked to the stochastic properties of the experiment and stochastic coherence must be preserved.

The assumption made in this work is that time is not an absolute concept, but it is related to an experiment, therefore each experiment has its own time structure.

This is a consequence of the observation that it is not possible to observe time evolution without the dynamic of information and information changes are relative to the experiment.

At the end of this section we will show that usual time is a special case of this definition of stochastic time. Indeed usual time is a law of chance and hence it has a deep link with the concept of proof, due to the fact that profs are laws of chance.

This rough idea of stochastic time for the case of a coin tossing experiment is given in figure 3.

Now let us give the complete description of the category $\mathcal{T}^X$.

For every observable $x \in |X|$ there is an object A_x of $\mathcal{T}^X$ called stochastic time point. A_x is the set of all σ_y that contain all the information of σ_x, i.e. $\sigma_y = \sigma_x$ and $x \subseteq y$. Intuitively A_x contains all the possible futures of x. if there is an arrow $f : A_y \to A_x$ then $A_y \subseteq A_x$. Therefore time arrows describe the process of information refinement.

Although useful in the understanding of stochastic time, the hypothesis of total information is too strong for a representation of the dynamic of information. For a mathematical description of chance it is more convincing the assumption of partial information.

Under this condition we do not know if an observable $\overline{x}$ belongs to the set of x s.t. $xTESTx_\infty$ or to the set of x s.t. $xSTATx_\infty$, hence we are forced to assume that $\mu_0 = \sigma_{\overline{x}}$. Thus an important aspect of chance is the dynamic of information, the acquisition of new information given by the observation of new outcomes, i.e. the process of refinement of information.

To understand how the flow of information works, assume that we have observed ht. In this information state, the possible future is given by all the arrows that come from a measure that gives the same probability to head and tail. If the successive outcome is h, then the new information state is A_{hth}, where we are forced to make a new hypothesis about the measure that governs the process (μ_0), i.e. the probability of h is $\frac{2}{3}$ and the probability of t is $\frac{1}{3}$.

This variability of information is a crucial point in the description of chance. Moreover note that this process goes in the opposite direction w.r.t. the arrows of $\mathcal{T}^X$.

From a mathematical point of view there is an elegant way to formalize the above properties. We can construct a new category $Sets^{\mathcal{T}^{X^{op}}}$ of objects that vary over the category of stochastic time.

More precisely the objects of $Sets^{\mathcal{T}^{X^{op}}}$ are contravariant functors from the category of stochastic time to the category of sets.

The functors are contravariant because, as we have noted, the process of information refinement reverses the arrows of time. This means that the dynamic of the information is described using the finite available information that goes from a finite past to the present. On the contrary the structure of stochastic time is defined using the important property that the meaning of every observable is given by its link with an infinite future where the sets STAT and TEST do not change.

All these ideas have the following mathematical representation. To each formula of the language of linear logic, or better to each experiment associated to the formulae, we can associate the category of contravariant functors from the category of stochastic time relative to the experiment to the category of sets.

This category, indeed a topos, is assumed as the mathematical description of the chance structure that generates the experiment, and it has interesting mathematical properties:

1) it is a complete set theory where sets vary over stochastic time

2) it has a powerful internal language of higher order, that means that it has formulae that speak over other formulae and it is possible to quantify over formulae, with a valid and complete proof system.

The category of chance is complex, but to give the intuition of how this structure can be used to show the soundness of the method it is sufficient to describe the object Ω that acts as a set of truth values for the internal language of the category of chance.

As a functor, Ω sends every stochastic time point A_x to the set of all cosieves on A_x. A cosieve S is a set of arrows s.t. if $f : A_y \rightarrow A_x$ is in S and $g : A_z \rightarrow A_y$ then $f \circ g \in S$. This means that if $f : A_y \rightarrow A_x$ belongs to S then every future of A_y belongs to S. Hence a cosieve on A_x describes a possible future that starts from x.

A formula α of the internal language is an arrow in the category of chance, i.e. a natural transformation, from a contravariant functor F to Ω ($\alpha : F \rightarrow \Omega$).

A formula α is $true$ in a time point A_x if its value on $F(A_x)$ is the maximal cosieve on A_x, i.e. the set of all arrows with codomain A_x. This means that α is $true$ in A_x iff it remains true in every future of A_x.

Therefore $true$ in A_x means stochastic-time invariant in A_x.

A formula is $true$ in the category of chance if it is $true$ in every stochastic time point, i.e. whatever will be the measure that governs the process.

Therefore we arrive at a mathematical description of law of chance: a law of chance is a $true$ formula in the category of chance.

It is possible to prove that every stochastic clique $a^{\phi}_{\overline{x}}$ can be represented in the category of chance by a suitable formula of the internal language $\alpha^{\phi}(\overline{x})$.

This representation is possible because in the internal language of the category of chance there is a formula $P(L) = r$ that is true in a time point A_x iff for every $\sigma \in A_x$ it holds that $\sigma(L) = r$. The term P is a categorical definition of measure. The definition is so abstract that it can represent different semantics of uncertainty. For instance if the category $\mathcal{T}$ is made of sets of possibility measures the proof system of the category of contravariant functors gives a valid and complete logic for possibilistic reasoning. The same holds for probabilities and upper and lower probabilities.

Using the term P, it is possible to show that every generic clique $a^{\phi}_{\overline{x}}$ can be represented in the internal language by a formula $\alpha^{\phi}(\overline{x})$.

This representation preserves the intended semantics. In fact it is possible to prove that $\alpha^{\phi}(\overline{x})$ is $true$ in A_x iff $\overline{x}$ and x are coherent observables.

Using this machinery we arrive at the proof of the soundness of the method. In fact, if π is a proof of a formula ϕ then for every observable $\overline{x}$ used to interpret π in the generic clique $a^{\phi}_{\overline{x}}$, it holds that $\alpha^{\phi}(\overline{x})$ is a $true$ formula, hence a law of chance.

This shows that proofs do not depend on the available information $\overline{x}$ neither on the unknown measure μ_0 that governs the process.

Let us analyze the proof $\vdash 1$, indeed an axiom of linear logic, to give a simple example of the method, and to discover from where usual time comes.

The atomic formula 1 is the witness of provability, in fact, in linear logic a formula is provable iff 1 belongs to its truth value.

1 is the neuter element w.r.t. the $\otimes$ connective, i.e. for every formula ϕ it holds that $1 \otimes \phi = \phi$.

Due to the stochastic meaning of the $\otimes$ connective the following equality must hold: $\mathcal{F}^{\epsilon(1)} \times \mathcal{F}^{\epsilon(\phi)} = \mathcal{F}^{\epsilon(\phi)}$. This means that for every n and for every $L \in \mathcal{F}^{\epsilon(1)}_n$, $M \in \mathcal{F}^{\epsilon(\phi)}_n$, it holds that $L \cap M = M$.

From this we obtain that for every $L \in \mathcal{F}^{\epsilon(1)}_n$ and every n, we have that $L = \Xi$.

This completely characterizes 1 as an experiment.

In fact as a stochastic process 1 is a family of constant random variables 1_n. Therefore 1 is an experiment that has only one possible outcome $1_0(\xi)$, that surely happens. An observable has the form $x = \{1_0(\xi), ..., 1_n(\xi)\}$. Moreover σ^1_x is the unique probability measure defined over $\{\Xi, \emptyset\}$. If π is the proof of 1, then π is interpreted in $a^1_{\overline{x}} = \{x : x \in |1| \wedge xSTAT_1\overline{x}\} = \{\{1_0(\xi)\}, \{1_0(\xi), 1_1(\xi)\}, ..., \{1_0(\xi), .., 1_n(\xi)\}, ..\}$. The formula $\alpha^1(\overline{x})$ that corresponds to $a^1_{\overline{x}}$ in the category of chance, is $P(\Xi) = 1$. It is clear that $\alpha^1(\overline{x})$ is a $true$ formula because for every probability measure σ it holds that $\sigma(\Xi) = 1$. Note that 1 is a law of chance: it does not vary in stochastic time, the outcome is always the same and it is invariant w.r.t. uncertainty, it does not depend on the measure that governs the process.

All these properties define the usual idea of time.

In fact a clock is nothing but an experiment that has a unique outcome: time passing, the result of a regular oscillation, that surely happens. An observable is a date, i.e. the number of outcomes observed. For instance today's date is nothing but the number of seconds passed after Christ's birth, i.e. the number of observed outcomes of our (abstract) clock.

Note that we have defined a semantics of logic directly based on the observables without the use of the concept of truth.

Truth can be seen as a law of chance, in fact it does not depend on the available observables. Unfortunately it is so abstract and mythological that it completely looses its connection with the observables and the computational aspects of logic (proofs). This is due to the fact that every proof is interpreted in a single element (i.e. true). For these reasons in logic there are two different semantics and proofs semantics is not a semantics for entire logic, for instance atomic formulae do not have an explicit interpretation (see [2]).

In the presented mathematical framework every formula has an interpretation (the category of chance constructed over the time structure of the experiment associated to the formula) rich enough to represent also the computational aspects of reasoning.

4 Conclusion

Let me conclude with a note on my personal interest in this research. I believe that the laws of chance can give an interesting contribute to answer the following question: why mathematics is reliable?

It is a common opinion that even a partial answer to this question could give some insight to the problem of the foundations of mathematics. There are many examples of the reliability of mathematics, for instance the existence of the planet Neptune has been foreseen only on the basis of mathematical computations.

We have seen that reasoning and computing, due to the strict link between proofs and typed λ-calculus, transforms sets of observables into set of observables in a sound mode. Where sound means that these transformations do not depend on the available data or on the (unknown) measure that governs the process, i.e. they are chance invariants.

My claim is therefore that mathematics is reliable because it is able to grasp some of these invariants that remain stable also in the presence of the high variability of outcomes due to randomness. This aspect gives us the possibility of defining non local rules (the one of logic and computations) used to give meaning to local observations (the one available to us), i.e. rules that allow us to forecast what we have not yet observed, like in the example of the discovery of the planet Neptune.

References

[1] R.J. Elliott. *Stochastic calculus and applications*. Springer–Verlag, 1982.

[2] J.-Y. Girard, Y. Lafont, and P. Taylor. *Proofs and Types*. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 1989.

[3] J.-Y. Girard. A new constructive logic: classical logic. *Mathematical Structures in Computer Science*, 1:255–296, 1991.

[4] J.-Y. Girard. On the unity of logic. *Annals of Pure and Applied Logic*, 59:201–217, 1993.

[5] P.R. Halmos. *Measure Theory*. Van Nostrand, 1950.

[6] C. Sossai. *Statistical reasoning Part 1: proofs and observables*. Submitted to IPMU 2008.

[7] C. Sossai. *Statistical reasoning Part 2: a stochastic semantics for logic*. Submitted to IPMU 2008.

[8] C. Sossai. *Statistical reasoning Part 3: proofs and laws of chance*. Submitted to IPMU 2008.

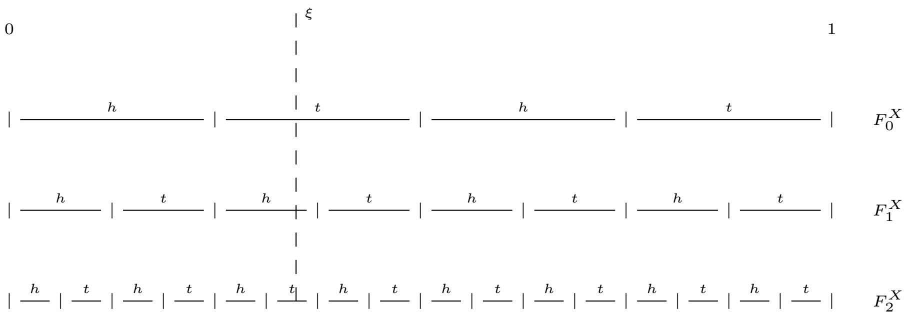

Figure 1: X: Fair coin experiment

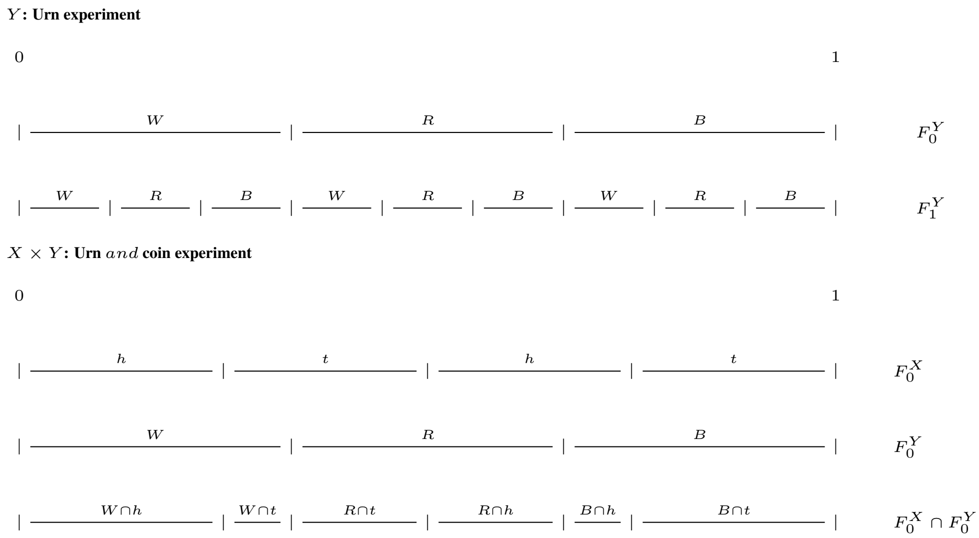

Figure 2: $X \times Y$: Urn and coin experiment

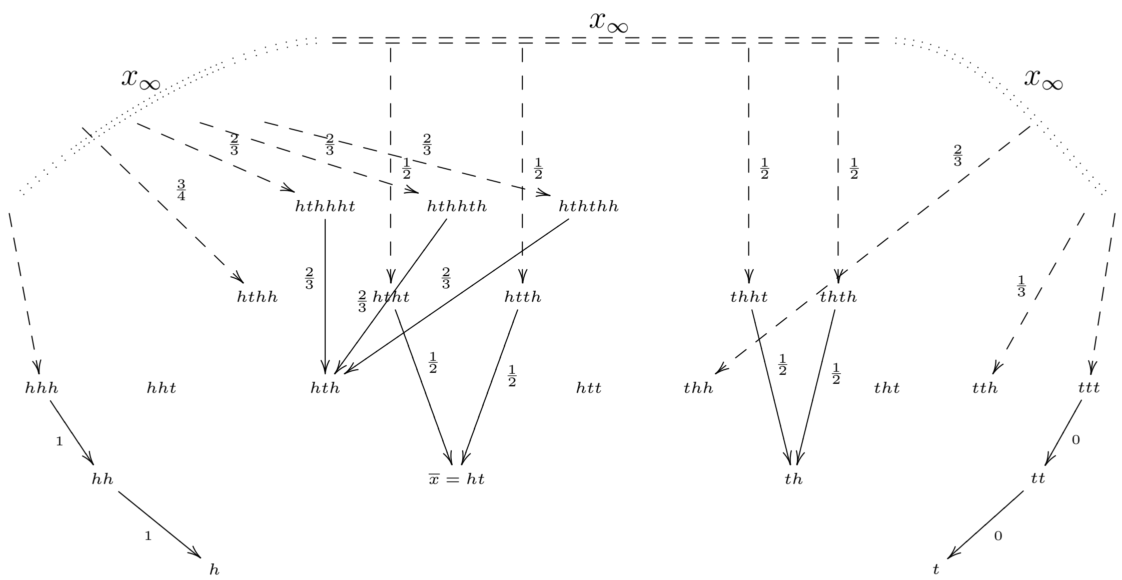

Figure 3: Sketch of the category of stochastic time (coin experiment)