Internal probability valuations

Tobias Fritz and David I. Spivak*

Categorical Probability and Statistics
June 7, 2020

Outline

What is the mathematical structure of story?

One might say, "every thought we have is a story snippet."

- Consider the events of your day; they form a story.
- When one discusses a game of Chess or Go, they tell a story.
- "Each protein vibrates at the frequency of the light it absorbs."

Let's invent something crude for story snippets and how they fit together.

- Our minds are organized to think in terms of time and space.
- We're interested in *what occurs* within this time and space.
- Abstracting, we're interested in *sheaves* on a *topological space* T.
  - A sheaf S is a world of story snippets, called *sections*.
  - Each story snippet occurs in a part $P \subseteq T$ of spacetime.

But what would be meant by characters?

Characters in a story-as-sheaf

A story character consists of two things.

- It has a body defining its expressive possibilities (internal and external).
- It has characteristic behaviors: some patterns are more likely.

We may think of the character's expressive possibilities as a sheaf.

- Like all sheaves, it's a world of story snippets, possibilities.
- You can make it interact with other characters through *relations*...
- ... but we'll ignore that unless someone asks about it.

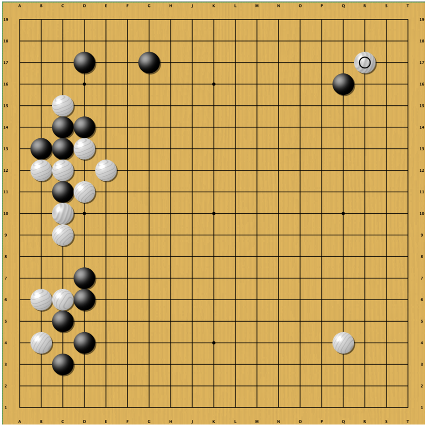
To have characteristic patterns, the body needs a notion of propensity.

- Propensity, tendency, likelihood, probability.
- We use our knowledge of tendencies and propensities to live.

In the context of sheaves, what should propensity be formalized as?

- This is the sort of thing people invented *stochastic processes* for.
- Today, we'll discuss a notion of propensity—*valuations*—on sheaves.

Running example: Go game



Basics of Go:

- The board has 19×19 intersections.
- Each gets black, white, or vacant.
- Board positions: $BP := 3^{19\times 19}$
- Some sequences in BP are legal.

What is overarching "spacetime" here?

- Go is a game of space: rectangles.
- We'll get precise on slide 8.

What is the "character's body" here?

- Body = legal sequences in BP.
- We'll get precise on slide 17.

What is the "propensity" here?

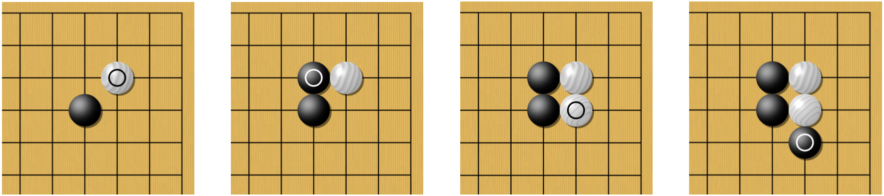
- Some sequences occur more often than others.
- How to formalize this? Results on slide 24.

Evaluation is local in Go

When we say "some sequences occur more than others", what do we mean?

- On the whole board $3^{19\times 19}$, you never see the same sequence twice.
- Instead, the statement refers to *local patterns*.

It is something like *this* that would be familiar to a go player:



Story: white invades, black attaches, white extends, black hits on the head.

- This is how go players talk. Note it is entirely local.
- We need to be able to evaluate the quality of local situations.
- How likely would this local sequence be in a professional game?

We'll get there, but we need to say a word about stochastic processes.

Stochastic processes

A. Khinchin invented stochastic processes (1930s) to model similar things.

- Behavior changing in time, and its characteristic propensities.
- For example, a random or a not-so-random walk.

The probability of rain in some region changes through time.

- In some sense it's random, but there are correlations across time.
- "What's the probability that it'll rain in Ottawa at second X?"
- You can give a value depending only on X (e.g. time of day, season).
- But how to encode that events at times t and $t+1$ are correlated?
- These correlations across time are what stochastic processes encode.

Example: probability distribution on functions $f : \mathbb{R} \longrightarrow \mathbb{R}$:

- $1/2 \cdot (\forall x. f(x) = x) + 1/2 \cdot (\forall x. f(x) = 0)$. Very correlated.
- $\forall x. [1/2 \cdot (f(x) = x) + 1/2 \cdot (f(x) = 0)]$. Totally uncorrelated.

Our approach

"Do not seek to follow in the footsteps of the wise; seek what they sought." *– Matsuo Bashō*

- We have what is probably a similar intention, but different approach.
- Stochastic process $\approx$ real-number random variables indexed by time.
- We want something more abstract, more general.
- We want a way to tell stories about more general characters.

Preview

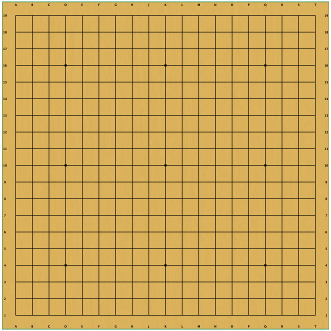
We define topological spaces and valuations internally to any topos $\mathrm{Shv}(T)$.

- A topological space $(X, \mathcal{O}_X)$ embodies some expressive possibilities.
- A valuation $\mu \colon \mathcal{O}_X \to [0,1]$ endows it with propensities.
- By doing this internally, these somehow vary over T.

To be satisfied, we want internally-defined valuations μ to:

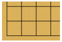
- Be internally sensible: understand how μ behaves logically.
- Be externally sensible: know what μ means sheaf-theoreatically.

Running example 1: space of Go rectangles



Consider the topological space $(Go, \mathcal{O}_{Go})$ with:

- $Go := \{r \subseteq 19 \times 19 \mid r \neq \varnothing,\ r \text{ rectangular}\}$

$$[A1..D3] = $$  $$\in Go$$

- Basic open: $U_r := \{r' \in Go \mid r' \subseteq r\}$
- Topology: generated by $(U_r)_{r \in Go}$.

Notes on the topology:

- Not T_1 because every open set containing r contains $r' \subseteq r$.
- Write $r \sqsubseteq r'$ if every open containing r also contains r'.
  - We have $r \sqsubseteq r' \Leftrightarrow r' \subseteq r$.
  - Say r is *less specific* than r', or that r' is a *specialization* of r.
- $(Go, \mathcal{O}_{Go})$ is a domain (continuous poset), if you know what that is.
- $U_r \cup U_{r'} \subset U_{r \cup r'}$, often strict.

Another example: Temporal type theory

This was our original motivation: probabilities on behavior types.

- Let $\mathbb{IR}$ be the space whose points are closed intervals $[a, b] \subseteq \mathbb{R}$.
- Basic open: $U_{(x_1,x_2)} := \{[a, b] \mid x_1 < a' \leq b' < x_2\}$.
- That is, take all $[a, b]$'s in the open interval (x_1, x_2).
- Nontrivial covers: $U_{(x_1,x_2)} = \bigcup_{x_1<y_1\leq y_2<y_2} U_{(y_1,y_2)}$.

A book called *Temporal type theory* is about a quotient of the above.

- Quotient by the translation action so that, e.g. $U_{(0,4)} = U_{(1,5)}$.
- This would add *stationarity*, but we won't discuss this today.

You might notice a similarity between $\mathbb{IR}$ and *Go*: an order on points.

- One may wonder, why not just take $\mathbb{R}$ as the space?
- You can do that!
- It's just that, as we'll see, valuations on $\mathbb{R}$ are not so interesting.

Plan

For the remainder of my time I'll:

- explain briefly what it means to work internally to a topos;
- internally define topological spaces, lower reals, and valuations; and
- discuss some mostly-proven theorems and conjectures.

Outline

Categories of sheaves

Given a topological space T, a *sheaf X on T* consists of

- A functor $X \colon \mathcal{O}(T)^{\mathrm{op}} \to \mathbf{Set}$, i.e.
  - To each open $U \subseteq T$, assign a set $X(U)$ of "*sections*".
  - To each $V \subseteq U$, assign a function $X(U) \to X(V)$, "*restriction*".
    - Given $x \in X(U)$, write $x\big|_V$ for its image in $X(V)$.
- Such that X satisfies the following "*Gluing condition*":
  - Suppose given opens with $U = \bigcup_{i \in I} V_i$ and sections $x_i \in X(V_i)$.
  - S'pose for all open $V \subseteq V_i \cap V_j$, have compatibility: $x_i\big|_V = x_j\big|_V$.
  - Then there exists a unique $x \in X(U)$ with $x\big|_{V_i} = x_i$ for all i.

A *morphism* of sheaves is just a natural transformation $\begin{tikzcd} \mathcal{O}_T \arrow[r, bend left, "X"] \arrow[r, bend right, "Y"'] \arrow[r, phantom, "\Downarrow\alpha"] & \mathbf{Set} \end{tikzcd}$.

Get a category $\mathrm{Shv}(T)$. It's a topos, but we'll get to that soon.

- Replacing "open" with "basic open" above, you get equivalent cat'y.
- For *Go*, we have $\mathrm{Shv}(Go) \cong \mathbf{Fun}\big((Go, \sqsubseteq), \mathbf{Set}\big)$.

Properties of a sheaf topos

A category $\mathrm{Shv}(T)$ of sheaves on a site has powerful reasoning capabilies.

- It has all small limits and colimits; denote terminal by 1.
- It is Cartesian closed, i.e. has internal homs.
- It has a natural numbers object $1 \xrightarrow{0} \mathbb{N} \circlearrowleft^{s}$; UMP = "induction".
- It has a subobject classifier, a notion of "truth values".

A *subobject classifier* is an object $\mathtt{Prop}$ and a map $1 \xrightarrow{\top} \mathtt{Prop}$.

- Note that any map out of 1 is vacuously monic.
- The pullback of a monic is monic, so given $f\colon B \to \mathtt{Prop}$, get monic

(just notation)
$$\begin{tikzcd} \{b : B \mid fb = \top\} \arrow[r, "!"] \arrow[d, hook] \arrow[dr, phantom, "\lrcorner", very near start] & 1 \arrow[d, hook, "\top"] \\ B \arrow[r, "f"'] & \mathtt{Prop} \end{tikzcd}$$

- Call $\top\colon 1 \to \mathtt{Prop}$ a *subobject classifier* if $\mathtt{Prop}^B \to \mathbf{Sub}(B)$ is iso.

In the topos **Set**, the subobject classifier is $1 \xrightarrow{\top} \{\top, \bot\}$.

Logic in a topos

Let $(T, \mathcal{O}_T)$ be a topological space. Let's consider $\mathtt{Prop} \in \mathsf{Shv}(T)$.

- It's the sheaf $U \mapsto \{U' \subseteq U\}$; restriction given by $U'\big|_V := V \cap U$.
- Logic: $\top, \bot, \wedge, \vee, \Rightarrow, \neg, \Leftrightarrow, \forall, \exists$, all about open subsets of T.
  - $\top$=biggest, $\bot$=smallest, $\wedge = \cap$, $\vee = \cup$,
  - $\exists(a : A).P(a) = \bigcup\{U \in \mathcal{O} \mid \exists a \in A(U) \mid P(a) = U\}$, etc.
  - All logical operations become operations on open subsets.
  - See Ingo Blechschmidt's thesis for a great introduction.

We can talk as though about ordinary sets, but "dog whistle" sheaves.[1]

- This is called "working internally" to the topos.
- This is what we'll do to define topological spaces and valuations.

---

[1] Dog-whistle politics is about language that sounds innocuous, but actually carries extra content for certain listeners. Here we're talking as though just about sets, but those with sensitive ears hear facts about sheaves.

Modalities: sublocales

One final point before we move on: modalities.

- A *modality* is a map $j \colon \mathtt{Prop} \to \mathtt{Prop}$ such that for all $P : \mathtt{Prop}$,
  - $P \Rightarrow jP$;
  - $jjP = jP$; and
  - $j(P \wedge Q) = jP \wedge jQ$ [equivalently, $(P \Rightarrow Q) \Rightarrow (jP \Rightarrow jQ)$]
- There is an equivalence between subtoposes and modalities.
  - For example, an open or closed subset of T, a point of T.
  - Each of these subspaces (more generally sublocales) has its own j.
  - Today's favorite modality: that for a point $t \in T$, denoted $@_t$.
- Adding j throughout a formula makes it about the j-subtopos.
  - E.g. adding j throughout $\forall(x : X).\exists(y : Y).P(x) \Rightarrow Q(y)$ yields
  - ... $j\forall(x : X).j\exists(y : Y).j(jP(x) \Rightarrow jQ(y))$.
  - If φ is the original formula, call the j-throughout version φ^j.
- Then φ^j holds in the big topos iff φ holds in the j-subtopos.
  - So $\varphi^{@_t}$ holds in $\mathrm{Shv}(T)$ iff φ holds at the point $t \in T$.

Outline

Internal topological spaces

To define a topological space internal to $\mathrm{Shv}(T)$, just write the definition.

- An *internal topological space* consists of an object $X \in \mathrm{Shv}(T)$...
- ...and a subobject $\mathcal{O}_X \subseteq \mathbf{Sub}(X)$ "open sets" satisfying some rules.
  - $X \in \mathcal{O}_X$;
  - $\varnothing \in \mathcal{O}_X$;
  - $\forall (U, V \in \mathcal{O}_X).(U \cap V) \in \mathcal{O}_X$; and
  - $\forall (S \subseteq \mathcal{O}_X).(\bigcup_{U \in S} U) \in \mathcal{O}_X$.
    - Logically, this would be $\forall (S \subseteq \mathcal{O}_X).(\exists (U : S).U) \in \mathcal{O}_X$.

Follow standard definition and use our sheaf-whistle. What does it mean?

Moerdijk's theorem

In *Spaced Spaces*, Moerdijk gives the semantics of the above construction.

- Every sheaf X on T has an associated bundle $\pi_X \colon \text{ét}(X) \to T$.
  - It's called the "étale space over T associated to X"
  - For any $i \colon U \subseteq T$, we have $X(U) \cong \{s \colon U \to \text{ét}(X) \mid \pi_X \circ s = i\}$.
- Moerdijk proves that a topology $\mathcal{O}_X$ on X as above is the same as...
- ... a coarsening of ét(X) over T, i.e. a diagram

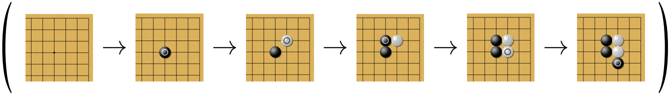
$$\begin{tikzcd} \text{ét}(X) \arrow[rr, "\text{id. on points}"] \arrow[dr, "\pi_X"'] & & \text{ét}(X, \mathcal{O}_X) \arrow[dl, "\pi_{(X,\mathcal{O}_X)}"] \\ & T & \end{tikzcd}$$

The opens of ét$(X, \mathcal{O}_X)$ correspond to elements of $\mathcal{O}_X$.

An internal space for Go

In Shv(Go), we can define a relevant internal space.

- Every spot is either black, white, or vacant $\{B, W, V\}$.
- Given a rectangle $r \subseteq 19 \times 19$, let U_r be the corresponding open.
- We can consider the sheaf $Pos(U_r) := \{f : r \longrightarrow \{B, W, V\}\}$.
- More interesting: the sheaf Seq of "legal sequences" of moves.
    - A list in Pos that starts empty, subsequent entries fill one vacancy.
    - Some subtlety on stone removal, but easy if you know the rules.



$\in Seq(O14..T19)$

- What topology should we put on Seq?
    - Possibility: order Seq by list prefix and use Alexandrov topology.
    - So a basic open is the set of futures of a given sequence.

Where we're going

We're going to define valuations μ internally, like we did for top'l spaces.

- A valuation will defined on an internal topological space $(X, \mathcal{O}_X)$.
- It'll be a map $\mu\colon \mathcal{O}_X \to [0,1]$ satisfying some conditions.
- But what exactly is $[0,1]$? Everything's supposed to be a sheaf on T.
    - Next we'll internally define $\underline{\mathbb{R}}$ = "lower reals".
    - Then we'll discuss their semantics.
    - The interval $[0,1]$ is a subsheaf of $\underline{\mathbb{R}}$.
- We'll finish this section by giving the definition of valuation.

Internal real numbers and their semantics

Dedekind (1872) defined real numbers in terms of "pairs of cuts" $r \subseteq \mathbb{Q}\times\mathbb{Q}$.

- His definition turns out to have very pleasing semantics in toposes.
  - Let $\mathbb{Q} \in \mathrm{Shv}(T)$ be the locally constant sheaf of rationals.
  - Define $R \subseteq \mathtt{Prop}^{\mathbb{Q}\times\mathbb{Q}}$ internally using Dedekind's axioms.
  - Dubuc (citing Joyal): R is isomorphic to sheaf of maps $T \longrightarrow \mathbb{R}$!
  - This is considered evidence that "internal language really works".
- We will use a similar construction called "lower reals": only one cut.

Internal definition of lower reals

Define $\underline{\mathbb{R}} \subseteq \mathtt{Prop}^{\mathbb{Q}}$ to be those $r_> : \mathbb{Q} \to \mathtt{Prop}$ such that:[2]

"nonempty"	$\exists(q : \mathbb{Q}).r_> q$
"down-closed"	$\forall(q, q' : \mathbb{Q}).(r_> q' \wedge q' > q) \Rightarrow r_> q$
"rounded"	$\forall(q : \mathbb{Q}).r_> q \Rightarrow \exists(q' : \mathbb{Q}).r_> q' \wedge q' > q$

The real is defined to be its set of rational lower-bounds.

- We'll write $q < r$ or $r > q$ from now on, rather than $r_> q$.
- For any rational q, can define $q_> \in \underline{\mathbb{R}}$ by $q_> q' := q > q'$.
- Write $r \leq r'$ to mean: $\forall q.(q < r) \Rightarrow (q < r')$.
- Define addition $(r_1 + r_2) \in \underline{\mathbb{R}}$ of lower reals $r_1, r_2 \in \underline{\mathbb{R}}$ by

$$q < (r_1 + r_2) \text{ iff } \exists(q_1, q_2).(q = q_1 + q_2) \wedge (q_1 < r_1) \wedge (q_2 < r_2).$$

- One can show $(\underline{\mathbb{R}}, 0, +)$ forms an ordered commutative monoid.

[2] Here $r_>$ is just a symbol; often people use $\delta : \mathbb{Q} \to \mathtt{Prop}$.

Semantics of lower reals

Semantically, the sheaf $\underline{\mathbb{R}}$ on T corresponds to:

- The sheaf that assigns to each $U \in \mathcal{O}_T$ the set

$$\underline{\mathbb{R}}(U) \cong \{f\colon U \to \mathbb{R} \mid f \text{ is lower semicontinuous}\}$$

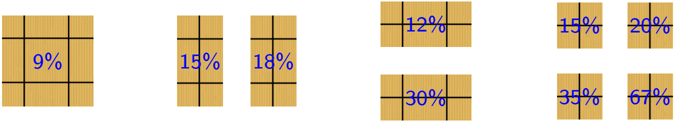
- Equivalently, we can say this in terms of a different topology on $\mathbb{R}$.
  - The sets $(U_r)_{r\in\mathbb{R}}$, with $U_r := \{r' \in \mathbb{R} \mid r<r'\}$, form a topology $\mathcal{U}$.
  - The sheaf $\underline{\mathbb{R}}$ is that of continuous maps to $(\mathbb{R}, \mathcal{U})$.

Let's return to our running examples to see what $\underline{\mathbb{R}}$ looks like there.

Lower reals in our examples

In the topos $\mathrm{Shv}(Go)$, a lower real assigns a lower s-c. function $f : Go \longrightarrow \mathbb{R}$.

- That is, to each rectangle r, a real number $f(r)$,...
- ... such that if $r' \subseteq r$ then $f(r') \geq f(r)$.



In the topos $\mathrm{Shv}(\mathbb{IR})$, a lower real assigns a lower s-c. function $f : \mathbb{IR} \longrightarrow \mathbb{R}$.

- That is, to each interval $[a, b]$, a real number $f[a, b]$,...
- ... such that if $[a', b'] \subseteq [a, b]$ then $f[a', b'] \geq f[a, b]$, and also...
- ... if $[a, b] = \bigcap\{[a_i, b_i] \mid i \in I\}$ then $f[a, b] = \sup_{i \in I} f[a_i, b_i]$.

Probability valuations

A *probability valuation* on $(X, \mathcal{O}_X)$ is a function $\mu \colon \mathcal{O}_X \to \underline{\mathbb{R}}$ such that:

(Strict)	$\mu(\bot) = 0$;
(Normalized)	$\mu(\top) = 1$;
(Monotonic)	$U \subseteq V \Rightarrow \mu(U) \leq \mu(V)$;
(Modular)	$\mu(U) + \mu(V) = \mu(U \cup V) + \mu(U \cap V)$; and
(Scott continuous)	for any directed subset $D \subseteq \mathcal{O}_X$, have

$$\sup_{U \in D} \mu(U) = \mu\Big(\bigcup_{U \in D} U\Big).$$

It's a reasonable notion of probability.

- On well-behaved spaces, valuations and probability measures coincide.
- It is also appealing from a CT point of view.

We can consider the above as an internal definition on $\mathrm{Shv}(T)$.

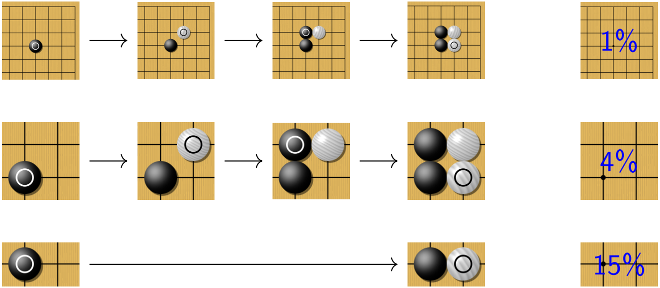
Go example

Recall that in the topos $\mathrm{Shv}(\mathit{Go})$, we had

- Seq of legal move-sequences (Alexandrov) as an internal space.
- Lower real = function $f\colon \mathit{Go} \to \mathbb{R}$ such that $r' \subseteq r \implies f(r') \geq f(r)$.

So what's a valuation $\mu\colon (\mathit{Seq}, \mathcal{O}_{\mathit{Seq}}) \to \underline{\mathbb{R}}$?

- It's a sheaf morphism, a natural transformation. Choose open U_r.
- An element $s \in \mathcal{O}_{\mathit{Seq}}(U_r)$ is a legal move prefix on rectangle r.
- It is assigned a value $\mu_{U_r}(s) \in \underline{\mathbb{R}}(U_r)$, a lower real.[3]



[3] All probabilities shown were casually made up; they're not from data.

Does it work as expected?

While a function $\mu \colon \mathcal{O}_X \to \underline{\mathbb{R}}$ looks good, does it work?

- Assigning a lower real to each legal Go sequence seems good.
- Implicitly it seems that we're imagining a valuation at each rectangle.
- Is that what we get when we evaluate the *semantics* of a valuation?

This leads us to the final section of the talk.

Outline

Choosing axioms

Under what conditions do valuations have “understandable” semantics?

- We choose some axioms on our topos; this is a bit of an art.
- Usually good to choose axioms that, as much as possible are:
  - small in number;
  - semantically meaningful;
  - logically simple; and
  - sufficient to prove something sensible.

Axioms on our topos

There is a type $\mathtt{Pt}$ and a function $@ : \mathtt{Pt} \longrightarrow \mathtt{Modality}$, such that

(Boolean "points") $\quad \forall (P : \mathtt{Prop})(t : \mathtt{Pt}). @_t P \vee (P \Rightarrow @_t \bot)$

(Enough "points") $\quad \forall (P : \mathtt{Prop}).\big(\forall (t : \mathtt{Pt}). @_t P\big) \Rightarrow P$

($\mathbb{N}$ "specialization-flabby") $\quad \forall (P : \mathbb{N} \longrightarrow \mathtt{Prop})(t : \mathtt{Pt}).$
$$@_t \exists (n : \mathbb{N}). P(n) \quad \Rightarrow \quad \exists (n : \mathbb{N}). @_t P(n).$$

Note that $\mathbb{Q}$ and $\mathbb{N}$ are internally bijective so Axiom 3 holds for $\mathbb{Q}$ too.

Families of valuations

Think of an internal space $(X, \mathcal{O}_X)$ as a space $\mathcal{X} \to T$ mapping to T.

- To every point $t \in T$, let $\mathcal{X}_t \subseteq \mathcal{X}$ denote the fiber subspace.
- If $t \sqsubseteq t'$, get continuous restriction map $\mathcal{X}_t \to \mathcal{X}_{t'}$.

A valuation $\mu \colon \mathcal{O}_X \to \underline{\mathbb{R}}$ induces a family of valuations $\mu_t \colon \mathcal{X}_t \to \mathbb{R}$.

- Internally, maps $@_t\mathcal{O}_X \to @_t\underline{\mathbb{R}}$ satisfying $@_t$-local def'n of valuation.
- This family $(\mu_t)_{t \in T}$ is compatible with restriction along $t \sqsubseteq t'$.
- It's also l-semicontinuous: for $U : \mathcal{O}_X$, $\mu(U) > q$ defines a lower real.

The above can be made precise internally and externally.

Valuations are @-locally defined

Assuming the three axioms above, we have a proof[4] of the following:

Theorem

Let $(X, \mathcal{O}_X)$ be an internal space. Restricting a valuation on X to a family of @-local valuations implements a bijection between

1. *valuations on X, and*
2. *those families of @-local valuations which are compatible and lower semicontinuous.*

This holds internally, and an analogous statement holds externally as well.

[4] The proof isn't fully complete, but we're fairly confident.

Upshot

Thus the internal notion of valuation on T has very reasonable semantics.

- It's a "measure space" at every point $t \in T$, such that
  - specialization preserves measure and
  - having probability strictly greater than q is an open condition.
- For Go, every rectangle gets a probability measure on legal sequences.
  - The likelihood of a sequence grows as the rectangle shrinks.
  - There's no continuity condition.
- Similarly for $\mathbb{IR}$, but there is a continuity condition in that case.

What should the poset of points be for stochastic processes?

- Reasonable options: finite subsets or closed intervals of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, etc.

While the conjecture also applies for the case of $\mathbb{R}$, it's a bit boring.

- The specialization order on points is trivial.
- So probability distributions at varying points are uncorrelated.

Outline

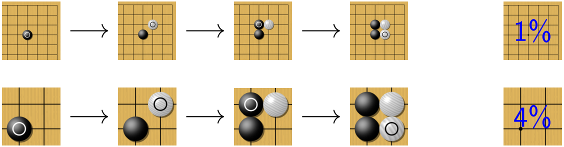
Summary

Valuations are an alternative to measures and have reasonable semantics.

- They are logical constructs and can be manipulated as such.
- In spatial (and more gen'l) toposes, they're defined locally at points:
  - The valuation is completely determined by its values on points.
  - A compatible, l-sc family of point valuations induces a valuation.
- We expect this formalism can treat all standard stochastic processes.

Main idea: to get correlations, use a poset of points.

- The probability of an event increases under restriction.



- Stories: usually mundane at a micro-level and novel in a larger context.

Questions and comments are welcome. Thanks!