Two layers of inference

Ryszard Paweł Kostecki

National Quantum Information Centre
Faculty of Mathematics, Physics, and Informatics, University of Gdańsk

23th Kraków Methodological Conference
8.11.19

Basic distinction

Two layers of logic: structural vs epistemic
(as stressed by Jaynes and Hintikka):

- **deductive** inference:
  - e.g. first order logic
  - premises: formulas that are truth valued (certain)
  - inference: turns certain premises to certain conclusions
- **inductive** inference:
  - e.g. probability theory + Bayes–Laplace rule
  - premises: formulas that are probability valued (plausible)
  - inference: turns plausible premises into most plausible conclusions

Can we extend Lawvere's views on logic with the above distinction?

- "In sheaf theory you do algebra vertically and topology horizontally".
- If: algebra $\leftrightarrow$ logic, and topology $\leftrightarrow$ logic, can the first be deductive, and other inductive?
- I will discuss the relevance of this structural claim in GR and QT, when approached via algebraico-geometric/category-theoretic perspective.

1. Toposes and general relativity

Toposes in a nutshell

- C: a small category
- $\mathbf{Set}^{C^{\mathrm{op}}}$: a category of presheaves over C
- J: a Grothendieck topology on C
- sheafification functor: $\mathbf{Set}^{C^{\mathrm{op}}} \to \mathbf{Sh}_J(C)$ = a Grothendieck topos
- $\mathcal{T}$: an elementary (Lawvere–Tierney) topos := a category that has all limits, exponentials, and a subobject classifier Ω.
- inner logic: higher order intuitionistic type theory
  Ω has a Heyting algebra structure
- one can formulate synthetic (structural) theories of algebraic, geometric, analytic,... objects in this type theory (or its subtheory), and then study various topos models
- there are many possible Lawvere–Tierney topologies j on $\mathcal{T}$, corresponding (if $\mathrm{sh}_j\mathcal{T}$ is representable as $\mathbf{Sh}_J(C)$) to different Grothendieck topologies (C, J)
- different sheafifications (and thus topos topology) can be seen as representing different epistemic criteria complementing the synthetic/structural side of a theory

Toposes in a nutshell

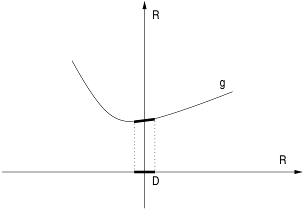
- Key insight: Yoneda lemma: one can embed C into $\mathbf{Set}^{C^{\mathrm{op}}}$ fully and faithfully by a family of functors $Hom_C(-, A) : C \to \mathbf{Set}^{C^{\mathrm{op}}}$, indexed by $A \in \mathrm{Ob}(C)$
- The type theory of a topos can be seen naively as a 'relative' set theory, with formulas $x \in_A B$ understood, in $\mathbf{Set}^{C^{\mathrm{op}}}$ as $x \in \mathrm{Nat}(Hom_C(A, -), Hom_C(B, -)) \cong \mathrm{Hom}_C(A, B)$.

SDG in a nutshell

- Synthetic theory of differential geometry, with intuitionistic proofs...
- ...represented in models that are sheafifications of $\mathbf{Set}^{C^\infty}$, where C^∞ is a category of (commutative) C^∞-algebras with smooth algebra homeomorphisms ("well adapted models")
- motivation #1: the spectral duality between the category *Man* of smooth manifolds and C^∞ has limitations:
  1) C^∞ is not cartesian closed, so the space of all smooth maps between two smooth manifolds is not a smooth manifold,
  2) pullbacks of smooth manifolds are not smooth manifolds (in general),
  3) jets and germs of smooth functions can be shown to be functors, but are not representable
- motivation #2: give exact sense to existence of infinitesimal objects s.t. $d \neq 0$ and $d^2 = 0$.

SDG in a nutshell

- start from a commutative unital ring R, $D := \{x \in R \mid x^2 = 0\} \subset R$, $D \neq \{0\}$
- Kock–Lawvere axiom: $\forall g : D \to R\ \exists! b : D \to R\ \forall d \in D\ \ g(d) = g(0) + d \cdot b$.



- Higher order Taylor series: $D_n := \{x \in R \mid x^{n+1} = 0\} \subset R$
- The space of k-jets in n variables:
  $D_k(n) := \{(x_1, \ldots, x_n) \in R^n \mid x_{i_1} \cdot \ldots \cdot x_{i_{k+1}} = 0\}$ for any k-tuple $(i_1, \ldots, i_{k+1})$
- Even more generally: Weil algebras:
  $R[X_1, \ldots, X_n]/(p_1(X_1, \ldots, X_n), \ldots, p_m(X_1, \ldots, X_n))$
- E.g. $D := Spec_R(R[X]/[X^2]) = \{d \in R \mid d^2 = 0\}$
- Kock–Lawvere: $\alpha : R[X]/(X^2) \ni (a, b) \mapsto [d \mapsto a + d \cdot b] \in R^D$ should be an isomorphism
- Generalised Kock–Lawvere axiom: for any Weil algebra W the R-algebra homomorphism $\alpha : W \to R^{Spec_R(W)}$ is an isomorphism.

SDG in a nutshell

- The generalised K–L axiom makes every jet representable.
- There are two other objects of infinitesimals:

$$\triangle := \{x \in R | \neg(x \in \mathrm{Inv}\ R)\} = \{x \in R | \neg\neg x = 0\}$$

$$\mathbb{\Delta} := \bigcup_{n>0} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{x \in R | \neg(x \# 0)\},$$

$$D(n) \subset D^n \subset D_k(n) \subset D_k^n \subset D(W) \subset \triangle^n \subset \mathbb{\Delta}^n.$$

where

$$x \# y := \exists n \in N \quad -\frac{1}{n} < x - y < \frac{1}{n}$$

$$\mathrm{Inv}\ R := \{x \in R | \exists y \in R \ \ xy = 1\}.$$

There are also some models in which one can consider also the object $\mathbb{I}$ of *invertible infinitesimals*:

$$\mathbb{I} := \{x \in R | x \in \mathbb{\Delta} \wedge x \in \mathrm{Inv}\ R\} = \bigcap_{n>0} (-\frac{1}{n}, \frac{1}{n}) - \{0\}.$$

Hence, $\mathbb{I} \subset \mathbb{\Delta} \supset \triangle$ and $\mathbb{I} \cap \triangle = \emptyset$.

SDG in a nutshell

An interpretation of these objects in well adapted models is:

smooth real line	$R = Y(\ell C^\infty(\mathbb{R})) = s(\mathbb{R})$
point	$x = Y(\ell(C^\infty(\mathbb{R})/(x))) = s(\{*\}) = \{x \in R \| x = 0\}$
first-order infinitesimals	$D = Y(\ell(C^\infty(\mathbb{R})/(x^2))) = \{x \in R \| x^2 = 0\}$
k$^{\text{th}}$-order infinitesimals	$D_k = Y(\ell(C^\infty(\mathbb{R})/(x^{k+1}))) = \{x \in R \| x^{k+1} = 0\}$
infinitesimals	$\Delta = Y(\ell C_0^\infty(\mathbb{R})) = \{x \in R \| \forall n \in N \quad -\frac{1}{n+1} < x < \frac{1}{n+1}\}$

The symbol Y denotes the Yoneda functor $\mathrm{Hom}(-, \ell A) =: Y(\ell A)$, while s denotes the functor $s : \mathbf{Man}^\infty \to \mathbf{Set}^{C^\infty}$

- When modelled in $\mathbf{Set}^{C^\infty}$, each point $x \in M$ becomes a functor that can be considered at different stages
- If modelled at the stage $\{*\}$, it corresponds to a global sheaf functor, and standard manifold picture, with no 'hidden structure' is restored.
- In order to have models of invertible infinitesimals (= Robinson's nonstandard analysis' infinitesimals), a forcing inside topos has to be made.
- It requires subsequently stronger conditions imposed on the underlying site (C, J) to have: jet representability, germ representability, germ representability with invertible infinitesimals.
- Local coordinate systems and implementation of rudimentary integration require to impose topological (not purely algebraic) conditions on Grothendieck topology: hence "measureability (epistemic potency of a theory) depends on the horizontal structure".

General relativity in a nutshell

- (M, g), where M is a smooth manifold,
- g is a bilinear symmetric 2-form field over M that has infinitesimally a form of a Minkowski metric $\mathrm{diag}(-1, +1, +1, +1)$
- $\nabla g = 0$ (Levi–Civita connection)
- Einstein equations are a set of PDEs: $R_{ij}(x) - \frac{1}{2} g_{ij}(x)(R(x) - 2\Lambda) = \kappa T_{ij}(x)$.
- General problem of observables in GR ("hole argument"): they are diffeomorphism invariant, so e.g. $g(x)$ has no observable meaning, and observables are essentially global.
- Fixing the global coordinate system = fixing the global gauge (there is still a local lorentzian gauge to be fixed in each point independently). But how?
- Bergmann–Komar ['56-'64] observables: a proposal to fix global coordinate system $(x_1, ..., x_4)$ as constructed from 4 linearly independent scalars constructed from g and its derivatives (more specifically, Weyl tensor). However, this does not work for spacetimes with symmetry.
- Typical approach: assume four global scalar fields $(\phi_1, \ldots, \phi_4)$ of some effectively noninteracting matter content and use it to define (and fix) $x_1(\phi_1, \ldots, \phi_4), \ldots, x_4(\phi_1, \ldots, \phi_4)$.

GR in SDG

- Guts'95,Guts–Grinkevich'96,...: formulation and investigation of the solutions
- Conceptual problem left open: what is the "physical meaning" of stages different from terminal (global sheaf)?
- RPK'05: Consider the stage as a specification of observer's context of observation, with different stages corresponding to different observers. Consider the stage dependence of points as a generalisation of a global gauge fixing $x_1(\phi_1, \ldots, \phi_4), \ldots, x_4(\phi_1, \ldots, \phi_4)$.
- Hence, we can think of the stage as fixing of the global frame of observation (diffeomorphism) performed in terms of the inner degrees of freedom of an observer, who substantialises the space in terms of his own subjective parameters, available to him at his stage.
- The terminal stage of $\{*\}$ is a global view.

Einstein algebras

- Geroch'72: Pass from M to $C^\infty(M)$, generalise the latter to a commutative $\mathbb{R}$-algebra, define g and ∇ on it, impose algebraic form of Einstein equations, consider it to be an algebraic generalisation of GR, and study representations.
- Heller'92:
  - Slightly different definition of Einstein algebra (a commutative $\mathbb{R}$-algebra C with a scalar product g on a module of C-derivations s.t. g has a lorentzian signature, $\nabla g = 0$, and algebraic form of Einstein equations holds).
  - Introduced Sikorski representation of Einstein algebras, substantially different from $C^\infty(M)$, and demonstrated that this allows to include space-time singularities of certain kind as a part of representation, substantiating the merits of Einstein-algebraic generalisation of GR.
  - Introduced sheafs of Einstein algebras over a Sikorski representation of a single Einstein/Lorentz algebra, which allows to include more general classes of singularities (studied in more details in Heller–Sasin'95).
- Lorentz algebra := Einstein algebra without assuming Einstein equations.
- Heller–Sasin'95 (implicitly) and Rosentock–Barrett–Weatherall'15 (explicitly): an Einstein/Lorentz algebra homomorphism is defined as an $\mathbb{R}$-algebra homomorphism ϕ s.t. the corresponding metric are equal under a push forward along ϕ.
- RPK'19: Given a category LA of Lorentz algebras, we can consider topos $\mathbf{Set}^{LA}$. Its sheafifications with respect to different subcanonical Grothendieck topologies will be called lorentzian toposes.

2. Information geometry and quantum theory

What are the state spaces?

- Pascal–Fermat '1654: probabilities, Huygens '1657: expectations
- XXth century probability theory = "measure theory with a soul" [M.Kac]: sets of (normalised) measures $p \in L_1(\mathcal{X}, \mu)$ or $p \in L_1(A)$, or integrals $p \in L$
- categorically equivalent formulations:
  (X, μ) - localisable measure spaces [Borel, Steinhaus, ..., Kolmorogov, Segal],
  A - maharanisable Dedekind complete boolean algebras [Carathéodory, Kappos],
  L - Banach preduals of proper abstract L_∞ spaces [Daniell, Riesz, Stone, Le Cam]
- in quantum theory one starts from 'pure' states ψ understood as vectors in a complex Hilbert space $\mathcal{H}$
- generally, a 'state' ρ = a linear operator on $\mathcal{H}$ that is trace class, i.e.
  $\mathcal{T}(\mathcal{H})^+ := \{\rho \in \mathfrak{B}(\mathcal{H}) \mid \mathrm{tr}_{\mathcal{H}}(|\rho|) < \infty, \rho \geq 0\}$, where:
  $\mathfrak{B}(\mathcal{H})$ = bounded linear operators on $\mathcal{H}$,
  (i.e., linear maps $x : \mathrm{dom}(x) \to \mathcal{H}$, s.t. $\mathrm{dom}(x) \subseteq \mathcal{H}$ and
  $\exists\lambda \in \mathbb{R}^+ \;\; \forall\psi \in \mathrm{dom}(x) \;\; \|x\psi\|_{\mathcal{H}} \leq \lambda\|\psi\|_{\mathcal{H}}$)),
  $\mathrm{tr}_{\mathcal{H}}(x)$ = trace of $x := \sum_i \langle\psi_i, x\psi_i\rangle_{\mathcal{H}} \in [0, \infty]$

Probability theory:

- Underlying structure: measure space $(\mathcal{X}, \mu)$
- Main spaces: **Probabilistic models:**

$$\mathcal{M}(\mathcal{X}, \mu) \subseteq L_1(\mathcal{X}, \mu)^+ := \{p : \mathcal{X} \to \mathbb{R} \mid \int_{\mathcal{X}} \mu |p| < \infty, p \geq 0\}$$

- e.g. Gaussian models: $\{p(\chi, (m, s)) = \frac{1}{\sqrt{2\pi s}} \mathrm{e}^{-\frac{(\chi - m)^2}{2s^2}} \mid (m, s) \in \Theta \subseteq \mathbb{R} \times \mathbb{R}^+\}$.
- Observables (estimators): functions $f : \mathcal{X} \to \mathbb{R}$

Quantum mechanics:

- Underlying structure: Hilbert space $\mathcal{H}$
- Main spaces: **Spaces of density matrices:**

$$\mathcal{M}(\mathcal{H}) \subseteq \mathcal{T}(\mathcal{H})^+ := \{\rho \in \mathfrak{B}(\mathcal{H}) \mid \mathrm{tr}_{\mathcal{H}}(|\rho|) < \infty, \rho \geq 0\}$$

- e.g. Gibbs states: $\{\mathrm{e}^{-\beta H} \mid \beta \in]0, \infty[\}$, for a fixed self-adjoint H.
- Observables: self-adjoint operators $x : \mathcal{H} \to \mathcal{H}$

W^*-algebras and integration

- A W^*-algebra $\mathcal{N}$:
  - a (noncommutative) algebra over $\mathbb{C}$ with unit $\mathbb{I}$,
  - with * operation s.t. $(xy)^* = y^*x^*$, $(x+y)^* = x^* + y^*$, $(x^*)^* = x$, $(\lambda x)^* = \lambda^* x^*$,
  - that is also a Banach space,
  - with $\cdot$, $+$, * continuous in the norm topology (implied by the condition $\|x^*x\| = \|x\|^2$),
  - such that there exists a Banach space $\mathcal{N}_\star$ satisfying the Banach space duality: $(\mathcal{N}_\star)^\star \cong \mathcal{N}$,
- Special cases:
  - if $\mathcal{N}$ is commutative
    then $\exists$ a measure space $(\mathcal{X}, \mu)$ s.t. $\mathcal{N} \cong L_\infty(\mathcal{X}, \mu)$ and $\mathcal{N}_\star \cong L_1(\mathcal{X}, \mu)$
  - if $\mathcal{N}$ is "type I factor"
    then $\exists$ a Hilbert space $\mathcal{H}$ s.t. $\mathcal{N} \cong \mathfrak{B}(\mathcal{H})$ and $\mathcal{N}_\star \cong \mathcal{T}(\mathcal{H})$.
- Hence, the element $\phi \in (\mathcal{N}_\star)^+$ provides a joint generalisation of probability density and of density operator.
- By means of embedding of $\mathcal{N}_\star$ into $\mathcal{N}^\star$, it is also an integral on $\mathcal{N}$.
- Hence, the subsets of $\mathcal{N}_\star^+$ can be considered as generic quantum state spaces.

Objects = spaces of noncommutative integrals

- Key fact: The above setting allows to develop full-fledged integration theory on noncommutative W^*-algebras, which generalises integration theory on measure spaces (with partial integration, conditional expectations, $L_p(\mathcal{N})$ spaces, etc...).
- Key fact #2: The noncommutative measure theory, focused on measures on the orthomodular lattices of projection operators (in any W*-algebra) is not equivalent, and has essentially less structure (e.g. it does not even allow to construct noncommutative L_p spaces).
- Hence, in noncommutative case Huygens wins with Fermat–Pascal: expectation/integral is more fundamental than probability measure.
- Key fact #3: Non-type I W^*-algebras are indispensable generalisation of the ordinary quantum mechanics in several important cases, e.g. the maximum entropy states in thermodynamical limit, required for an exact derivation of Hawking and Unruh effects.

Troubles with deduction and spectrality

Quantum logic: [von Neumann'32, Birkhoff–von Neumann'36, Mackey'57, Gleason'57, Piron'64, Kochen–Specker'65,...]

- main structure: orthomodular orthocomplemented posets or lattices
- main advantage: direct generalisation of the structure of Borel–Steinhaus–Kolmogorov probability theory
- main disadvantages:
  - no-go for tensor products [Randall–Foulis'81, Aerts'81]
  - noncommutative integration theory is strictly more general
  - several internal ambiguities (semantic: see Redei'01; structural: e.g. at least 5 different negation hooks)

Spectral paradigm:

- no direct generalisation of algebraic spectral duality (Gel'fand) to noncommutative case
- noncommutative integration perspective: expectations $\gg$ measures/pure states
- operational/convex theoretic perspective: estimation of density matrix ρ is based on probability measures $\mathrm{tr}_{\mathcal{H}}(\rho E(\mathcal{X},\mu))$, where $E(\mathcal{X},\mu)$ is semi-spectral measure, that does not determine uniquely any projectors

Conclusion: a shift from deductive quantum logic of projectors/measures to quantum inductive inference of noncommutative integrals.

Topos theoretic approach

- Isham–Butterfield'98+, Isham–Doering'06+,...: topos of presheaves over commutative subalgebras of a fixed W^*-algebra, with inclusions as morphisms
- main advantage: allows to show that the Kochen–Specker theorem is equivalent to nonexistence of a global cross section of a spectral presheaf in such topos
- allows to rephrase *some* of the structure of quantum theory inside topos, in a such a way that Heyting algebra structure gains more fundamental role than quantum logics (OML's/OMP's)
- yet, all of the the key troubles of quantum logic and spectral paradigm remain:
  - topos is a cartesian closed category, while $\otimes$ of quantum theory is noncartesian monoidal $\Rightarrow$ no account for naturality of tensorial structure (and thus such things like entanglement, teleportation,...)

    (this is addressed by an alternative approach by Abramsky–Coecke'04+: taking symmetric monoidal category as a departure point (valid only in finite dimensions), with no topos structure and no quantum logic)

  - algebraic geometry of topos is commutative, while generic quantum probability theory is a noncommutative integration theory

What are the morphisms?

- commutative probability theory:
  Bayes'1763–Laplace'1774 rule
  Kolmogorov'1933: conditional expectations various people soon after:
  markovian (= normalised positive linear) maps
- quantum theory:
  von Neumann'1932–Lüders'55 'projective state reduction' rule
  Moy–Nakamura–Turumaru'54: conditional expectations
  Stinespring'55: completely positive maps ("quantum markovian")
- all those mappings can be viewed as inductive inference, e.g. change state due to change of information
- first categorical formulation with markov maps explicitly considered as categorical morphisms: Chencov'1965 & Morse–Sacksteder'1966.

Chencov's programme of categorical geometrostatistics

- Independent discoveries that 'Fisher information matrix' is a riemannian metric tensor on the space of probabilities (a.k.a. "statistical manifold"): Hotelling'29 (unpublished), Rao'45, Jeffreys'46.
- Chencov'64: introduced an affine connection on statistical manifold
- Chencov'65: paper "Categories of mathematical statistics" (as a reference for category theory: Russian '61 translation of Godement's 'Topologie algébrique et théorie des faisceaux'!)
- Chencov'68: generalised pythagorean theorem for relative entropies
- Chencov'69: characterisation of all riemannian–affine geometries that are monotone under markovian morphisms
- Morozova–Chencov'85,'89, Petz'94: characterisation of riemannian geometries of quantum state spaces that are monotone under quantum markovian morphisms
- Jenčová'03: characterisaton of monotone quantum affine connections

Chencov's programme of categorical geometrostatistics

Chencov'72: monography summarising '64-'72 work.

On the first page of introduction:

"The system of all statistical decision rules of all thinkable statistical problems taken together with a natural operation of composition forms an algebraic category. This category gives birth to a homogeneous geometry of families of probabilistic laws, in which the families play the role of 'figures', while decision laws describe 'movements'. Two families are congruent if and only if, when they are having the same statistical properties. The subject of this monography most exactly could be described by a notion 'geometrostatistics'."

Main 'postchencovian' plot twists in information geometry

- CP maps are limited as morphisms, because they pressume no correlations (e.g., a tensor product) contained in initial states of the global system, which is the opposite of the generic case in algebraic QFT [Reeh–Schlieder'61]
- Various probabilistic updating rules (Bayes–Laplace, Jeffrey,...) are special cases of constrained relative entropy maximisation [Williams'80,...]
- The same is for von Neumann–Lüders rule [Hellmann–Kamiński–RPK'14] and partial trace [Munk-Nielsen'15], which together accounts for everything in CP maps that is neither an unitary evolution nor a tensor product (thus, the 'purely inferential' layer).
- A Taylor expansion of large class of relative entropies induces a riemannian metric and two torsion-free affine connections [Ingarden et al'82, Eguchi'83,...]
- In particular, for markov monotone [Csiszár'63–Morimoto'63, Kosaki'82–Petz'85] relative entropies one regains precisely Chencov-type geometries, both in probabilistic and quantum case [Eguchi'83, Lesniewski–Ruskai'99]
- It turned out that information geometry has essentially two different layers: one associated with monotonicity under markov maps (as above), another associated with generalised pythagorean theorem: Brègman ['67] relative entropies, and resulting dually flat/hessian geometries.
- Structural conclusion: Start from Brègman relative entropies, with entropic projections as morphisms, and their Taylor expansion as local geometric structure.

Quantum information divergences/relative negentropies

Quantum information divergence $D : \mathcal{M}(\mathcal{H}) \times \mathcal{M}(\mathcal{H}) \to [0, \infty]$
s.t. $D(\rho, \sigma) = 0 \iff \rho = \sigma$.

E.g.

- $D_1(\rho, \sigma) := \mathrm{tr}_{\mathcal{H}}(\rho \log \rho - \rho \log \sigma)$ [Umegaki'62]
- $D_{1/2}(\rho, \sigma) := 2\left\|\sqrt{\rho} - \sqrt{\sigma}\right\|^2_{\mathfrak{G}_2(\mathcal{H})} = 4\mathrm{tr}_{\mathcal{H}}(\frac{1}{2}\rho + \frac{1}{2}\sigma - \sqrt{\rho}\sqrt{\sigma})$
  (Hilbert–Schmidt norm²)
- $D_{L_1(\mathcal{N})}(\rho, \sigma) := \frac{1}{2}\|\rho - \sigma\|_{\mathcal{T}(\mathcal{H})} = \frac{1}{2}\mathrm{tr}_{\mathcal{H}}|\rho - \sigma|$ (L_1/trace norm)
- $D_\gamma(\rho, \sigma) := \frac{1}{\gamma(1-\gamma)}\mathrm{tr}_{\mathcal{H}}(\gamma\rho + (1-\gamma)\sigma - \rho^\gamma\sigma^{1-\gamma})$; $\gamma \in \mathbb{R} \setminus \{0, 1\}$
  [Hasegawa'93]
- $D_{\alpha,z}(\rho, \sigma) := \frac{1}{1-\alpha} \log \mathrm{tr}_{\mathcal{H}}(\rho^{\alpha/z}\sigma^{(1-\alpha)/z})^z$; $\alpha, z \in \mathbb{R}$
  [Audenauert–Datta'14]
- $D_{\mathfrak{f}}(\rho, \sigma) := \mathrm{tr}_{\mathcal{H}}(\sqrt{\rho}\,\mathfrak{f}(\mathfrak{L}_\rho\mathfrak{R}_\sigma^{-1})\sqrt{\rho})$; $\mathfrak{f}$ operator convex, $\mathfrak{f}(1) = 0$
  [Kosaki'82,Petz'85]

for $\mathrm{ran}(\rho) \subseteq \mathrm{ran}(\sigma)$, and with all $D(\rho, \sigma) := +\infty$ otherwise.

Entropic paradigm: absolute and relative

- Gibbs'1902, Elsasser'37, Jaynes'57, Ingarden–Urbanik'62,...:

  constrained maximisation of **absolute** entropy
  (e.g., $S(\rho) = -D(\rho, \psi)$ with a fixed prior $\psi = \mathbb{I}/\dim \mathcal{H}$)
  as a method of **model construction**:

  $$\rho(\mathrm{constraints}) := \arg\sup\{S(\omega) \mid \mathrm{constraints}(\omega)\}$$

  selecting a specific class of models $\mathcal{M}$ with elements parametrised by allowed values of constraints' parameters and maximally noninformative with respective to anything else

- Kullback'59, Good'63, Hobson'69,...:

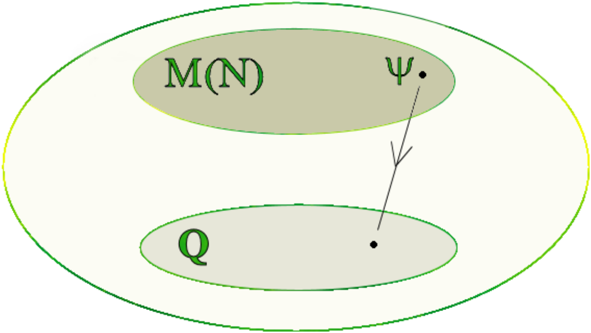
  minimisation of $D(\rho, \psi)$ as a method of state transformation (estimation, learning, updating,...) from ψ onto a set that satisfies given constraints.

Quantum entropic projections

Let $\mathcal{Q} \subseteq \mathcal{T}(\mathcal{H})^{+}$ be such that
for each $\psi \in \mathcal{M}(\mathcal{H})$
there exists a unique solution

$$\boxed{\mathfrak{P}^{D}_{\mathcal{Q}}(\psi) := \arg\inf_{\rho\in\mathcal{Q}} \{D(\rho,\psi)\}\,.}$$

It will be called an entropic projection.



Quantum measurement, bayesianity, and maximum relative entropy

- Williams'80, Warmuth'05, Caticha&Giffin'06:

  the Bayes–Laplace rule:

$$p(x) \mapsto p_{\text{new}}(x) := \frac{p(x)p(b|x)}{p(b)}.$$

  is a special case of

$$p(\chi) \mapsto p_{\text{new}}(\chi) := \arg\inf_{q \in \mathcal{Q}} \{D_1(q,p)\}\,; \quad D_1(q,p) := \int_{\mathcal{X}} \mu(\chi) q(\chi) \log\left(\frac{q(\chi)}{p(\chi)}\right).$$

- Douven&Romeijn'12: the Bayes–Laplace rule is also a special case of $p \mapsto \arg\inf_{q \in \mathcal{Q}} \{D_1(p,q)\} = \mathfrak{P}^{D_0}_{\mathcal{Q}}(p)$, where $D_0(p,q) = D_1(q,p)$.
- Lüders' rules [Lüders'55]:

$$\rho \mapsto \rho_{\text{new}} := \sum_i P_i \rho P_i \;\; (\text{'weak'}) \qquad \rho \mapsto \rho_{\text{new}} := \frac{P\rho P}{\mathrm{tr}_{\mathcal{H}}(P\rho)} \;\; (\text{'strong'})$$

- Bub'77'79, Caves–Fuchs–Schack'01, Fuchs'02, Jacobs'02:

  Lüders' rules should be considered as rules of inference (conditioning) that are quantum analogues of the Bayes–Laplace rule

Quantum bayesian inference from quantum entropic projections

- RPK'13'14, F.Hellmann–W.Kamiński–RPK'14:

  1. weak Lüders' rule is a special case of $\rho \mapsto \arg\inf_{\sigma\in\mathcal{Q}} \{D_1(\rho,\sigma)\}$ with
  $$\mathcal{Q} = \{\sigma \in \mathcal{T}(\mathcal{H})^+ \mid [P_i,\sigma] = 0\ \forall i\}$$
  2. strong Lüders' rule derived from $\rho \mapsto \arg\inf_{\sigma\in\mathcal{Q}} \{D_1(\rho,\sigma)\}$ with
  $$\mathcal{Q} = \{\sigma \in \mathcal{T}(\mathcal{H})^+ \mid [P_i,\sigma] = 0,\ \mathrm{tr}_{\mathcal{H}}(\sigma P_i) = p_i\ \forall i\}$$
  under the limit $p_2,\ldots,p_n \to 0$.
  3. hence, weak and strong Lüders' rules are special cases of quantum entropic projection $\mathfrak{P}^{D_0}_{\mathcal{Q}}$ based on relative entropy $D_0(\sigma,\rho) = D_1(\rho,\sigma)$.

Bayes–Laplace and Lüders' conditionings are special cases of entropic projections
$\Rightarrow$ "quantum bayesianism $\subseteq$ quantum relative entropism".

Quantum measurements from quantum entropic projections

- Hence: the rule of maximisation of relative entropy (entropic projection on the subspace of constraints) can be considered as a nonlinear generalisation of the dynamics describing *elementary* "quantum measurement".
- F.Hellmann–W.Kamiński–RPK'14: also quantum analogue of Jeffreys' rule follows
- M.Munk–Nielsen'15: partial trace is also entropic projection (for strictly positive states)
- more measurements and more general results: RPK&M.Munk–Nielsen'19 (under construction)
- these results are for D_0 and/or D_1; however there are many more D...
- how general measurements can be derived from entropic projections, allowing both D and Q to vary?

Generalised pythagorean equation

- The choice of the set $\mathcal{Q}$ for which the entropic projection $\mathfrak{P}^D_{\mathcal{Q}}$ exists and is unique depends very strongly on the structure of D: the choice of principle of inference (D) determines the accepted data types ($\mathcal{Q}$).
- We need some principle constraining D that would guarantee existence, uniqueness, and good composition properties of D-projections.
- We say that D satisfies a generalised pythagorean equation at $\mathcal{Q}$ iff [Chencov'68]

$$D(\phi,\psi) = D(\phi,\mathfrak{P}^D_{\mathcal{Q}}(\psi)) + D(\mathfrak{P}^D_{\mathcal{Q}}(\psi),\psi) \;\; \forall(\phi,\psi)\in\mathcal{Q}\times\mathcal{M}.$$

- Thus, information divergence decomposes additively under a projection onto a suitable subspace, hence we have a nonlinear, yet additive (!), decomposition: **data = signal + noise**
- Example 1: If $\mathcal{Q}$ forms an affine subset of $\mathfrak{G}_2(\mathcal{H})^+$ under $\rho\mapsto\sqrt{\rho}$, then:

$$\left\| x - \mathfrak{P}^{D_{1/2}}_{\mathcal{Q}}(z)\right\|^2_{\mathfrak{G}_2(\mathcal{H})} + \left\|\mathfrak{P}^{D_{1/2}}_{\mathcal{Q}}(z) - z\right\|^2_{\mathfrak{G}_2(\mathcal{H})} = \|x-z\|^2_{\mathfrak{G}_2(\mathcal{H})}.$$

- Example 2: If $\mathcal{Q} := \{\phi\in\mathfrak{G}_1(\mathcal{H})^+_1 \mid \phi(h) = \mathrm{const}\}$, then [Donald'90]

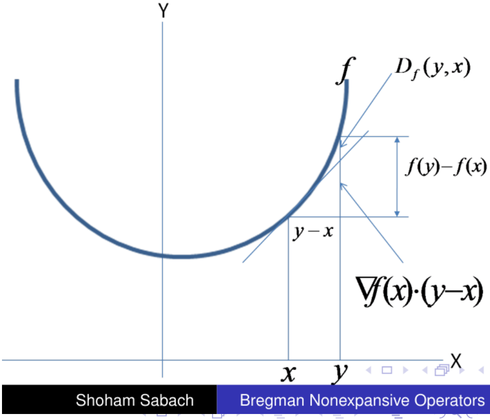
$$D_1(\phi,\psi^h) + D_1(\psi^h,\psi) = D_1(\phi,\psi) \;\; \forall(\phi,\psi)\in\mathcal{Q}\times\mathfrak{G}_1(\mathcal{H})^+_1.$$

Brègman relative negentropies D_Ψ

Brègman'67:
Let $f : \mathbb{R}^n \to]-\infty, \infty]$ be convex and proper ($\mathrm{efd}(f) := \{x \in \mathbb{R}^n \mid f(x) \neq \infty\} \neq \varnothing$). Then:

$$D_f(y,x) := f(y) - f(x) - \sum_{i=1}^{n} (y-x)_i [(\mathrm{grad} f)(x)]^i$$

- Jones–Byrne'90: D_f is characterised by the generalised pythagorean equation
- Bauschke–Borwein–Combettes'01: generalisation of D_f from $\mathbb{R}^n$ to arbitrary reflexive Banach space X under some additional conditions on f
- RPK'17: generalisation to $\tilde{D}_f$ defined over probabilistic, quantum, JBW, and other more general state spaces M via $\tilde{D}_f := D_f(\ell(\cdot), \ell(\cdot))$, where $\ell : M \to X$ is generally a nonlinear map



Categories of brègmannian entropic projections

- $\mathtt{Cvx}(\ell, f)$:
  - objects: ℓ-closed ℓ-convex subsets of M, including empty set
  - morphisms: $\mathfrak{P}^{\tilde{D}_f}_{Q}$, including empty arrows
  - composition: $\mathfrak{P}^{\tilde{D}_f}_{Q_2} \circ \mathfrak{P}^{\tilde{D}_f}_{Q_1} = \mathfrak{P}^{\tilde{D}_f}_{Q_1 \cap Q_2}$
- $\mathtt{Aff}(\ell, f)$: as above, but Q restricted to ℓ-affine ℓ-closed sets: the category of generalised pythagorean theorem
- $\mathtt{Cvx}^{\subseteq}(\ell, f)$, $\mathtt{Aff}^{\subseteq}(\ell, f)$: as two above, respectively, but with composition rule restricted to $Q_2 \subseteq Q_1$ (inclusion of convex/affine sets, in some analogy to Isham–Butterfield inclusion of commutative algebras)
- RPK'17: specific examples of above categories (in particular: a class of categories associated naturally with noncommutative Orlicz spaces over semi-finite W*-algebras and nonassociative L_p spaces over semi-finite JBW-algebras)
- The above categories provide the foundation of the nonmarkovian version of categorical geometrostatistics.
- Two natural directions to follow:
  1) semantics via adjunctions and monads/comonads
  2) localisation via toposes

Adjointness in (deductive) foundations

- Lawvere'63,'69:
  - $\mathcal{C}$: a category of deductive systems:
  - objects: formulas,
  - arrows: proofs/deductions.
  - $\mathcal{D}$: a category of geometric structures
  -

$$
\begin{tikzcd}
\mathcal{D} \arrow[d, bend left, "\substack{\texttt{syntax} \\ (\textit{formalisation})}"] \\
\mathcal{C} \arrow[u, bend left, "\substack{\texttt{semantics} \\ (\textit{interpretation}/\textit{model})}"] \arrow[u, phantom, "\dashv"]
\end{tikzcd}
$$

- examples:
  - $\mathcal{C}$:= typed λ-calculi with surjective pairing, $\mathcal{D}$:= category of cartesian closed categories; $\mathcal{C} \cong \mathcal{D}$ (Lambek'68,...)
  - $\mathcal{C}$:= extensional Martin-Löf theories, $\mathcal{D}$:= category of locally cartesian closed categories; $\mathcal{C} \cong \mathcal{D}$ (Seely'84,...)
  - $\mathcal{C}$:= intuitionistic higher order type theories, $\mathcal{D}$:= category of toposes with canonical subobjects and strict logical morphisms preserving canonical subobjects; adjointness (Lambek'74, Volger'75, Fourman'77,...)

Early categorical settings for inductive inference

- Chencov'65, Morse–Sacksteder'66:
  - objects: spaces of probability densities (subsets of L_1 spaces)
  - morphisms: Markov (i.e. linear, positive, and normalisation preserving) maps
- quantum generalisation (implicit: many authors in late 60s/early 70s):
  - objects: spaces of density matrices/normal states on W^*-algebras (subsets of noncommutative L_1 spaces)
  - morphisms: completely positive trace preserving linear maps
- in both probabilistic and quantum case this setting was used by Chencov and others to characterise such classes of geometric structures (riemannian metrics, affine connections) on objects of these categories that are monotonically decreasing under morphisms
- important observations:
  1. inductive inference categories are inherently **geometric**, with geometric properties encoding specific prescription of 'optimal' methods of model construction and inductive inference ("Jaynes–Chencov principle")
  2. for each specific method/category of inductive inference, there are different optimal **experimental designs** that can be analysed with it (e.g. χ^2 test makes no sense for a small sample size, the Bayes–Laplace rule is inapplicable to data given by average values, etc...)

Epistemic adjointness [RPK'12,'16]

- **Postulate, ver.1**: given the category $\mathtt{IndInf}$ of inductive inferences, the optimal category $\mathtt{ExpDes}$ of experimental designs corresponding to $\mathtt{IndInf}$ should be such that there exist two adjoint functors:

$$
\begin{tikzcd}
\mathtt{IndInf} \arrow[d, bend left, "\overset{\texttt{syntax}}{(\textit{predictive verification})}"] \arrow[d, phantom, "\dashv"] \\
\mathtt{ExpDes} \arrow[u, bend left, "\overset{\texttt{semantics}}{(\textit{model construction})}"]
\end{tikzcd}
$$

  i.e., the method of model construction should be the most effective solution of the problem provided by the given predictive verification.
- **Postulate, ver.2**: Given the category $\mathtt{IndInf}$, the admissible family of possible experimental design categories and the corresponding adjoint functors should be given by specifying a **comonad** on $\mathtt{IndInf}$.
- Dually, given $\mathtt{ExpDes}$, a monad on it describes a range of admissible inductive inference settings applicable optimally to it.

Resource theoretic perspective [RPK'16]

- **Embedding**:= a full and faithful functor $F : \mathcal{C} \to \mathcal{D}$.
  - it is **extensive** iff $F(\mathcal{C})$ is a subcategory of $\mathcal{D}$
  - it is **intensive** iff $\exists! G$ such that $F \dashv G$ and the unit of adjunction is a natural isomorphism.
  - an intensive embedding can be seen as a translation from more coarse-grained/concrete to more refined/abstract description
  - it gives rise to a comonad E on $\mathcal{D}$
- Let us also introduce a monad T on $\mathcal{D}$, representing the **allowed (free) operations** on $\mathcal{D}$.
- Assuming that $\mathcal{D}$ is equipped with a terminal object $\mathbf{1}$, an object x in $\mathcal{D}$ will be called a **free resource** iff $\exists$ an element f of T such that $f : \mathbf{1} \to x$.
- Taking $\mathcal{D}$ to be given by an inductive inference category $\mathtt{IndInf}$, we define a **categorical resource theory** as a triple $(\mathtt{IndInf}, E, T)$, where:
  - epistemic comonad E on $\mathtt{IndInf}$ provides specification of compatible experimental designs, via the corresponding syntax/semantics adjointness (the choice between Eilenberg–Moore and Kleisli constructions in this case depends on whether one wants to be maximally restrictive or maximally inclusive w.r.t. the range of admitted $\mathtt{ExpDes}$)
  - action monad T on $\mathtt{IndInf}$ provides specification of free operations and free resources
- one can consider a lax morphism of free operations monad T from inductive inference category $\mathcal{D}$ to experimental design category $\mathcal{C}$ along the (nonunique) right adjoint functor representing the experimental design comonad E

Entropic model construction as a functor [RPK'16,'17,'19]

- $\mathtt{Cvx}(\ell, f)$ as a model of $\mathtt{IndInf}$
- $\mathtt{ExpDes}$: a category with data sets of configuration parameters as objects, with arrows given by data sets describing registration parameters of experimental operations, and composition representing allowed compositions of experimental operations.
- F :$\mathtt{ExpDes}\rightarrow\mathtt{IndInf}$ provides an idealisation of finite data sets: mapping sets in $\mathrm{Ob}(\mathtt{ExpDes})$ into the ℓ-closures of their ℓ-convex envelopes, and mapping sets in $\mathrm{Arr}(\mathtt{ExpDes})$ into entropic projections onto ℓ-closures of the ℓ-convex envelopes of these sets.
- The adjoint functor: given by forgetting everything except the convex sets used as constraints.
- Taken together, they determine and epistemic comonad E on $\mathtt{Cvx}(\ell, f)$.
- The action monad on $\mathtt{Cvx}(\ell, f)$ is specified differently, using so-called Brègman monotone operations, which provide an implementation of Mielnik's ['69,'73] idea of nonlinear transmitters [details: RPK'17]

Smooth quantum information geometries

Taylor expansion of D induces a generalisation of a riemannian geometry on $\mathcal{M}(\mathcal{N})$.

- $\mathcal{M}(\mathcal{H}) := \{\rho(\theta) \in \mathcal{T}(\mathcal{H}) \mid \rho(\theta) > 0, \theta \in \Theta \subseteq \mathbb{R}^n \text{ open}, \theta \mapsto \rho(\theta) \text{ smooth}\}$ is a C^∞-manifold
- Jenčová'05: a general construction of smooth manifold structure on the space of all strictly positive states over arbitrary W^*-algebra, with tangent spaces given by noncommutative Orlicz spaces.
- Eguchi'83/Ingarden et al'82/Lesniewski–Ruskai'99/Jenčová'04:
  Every smooth divergence D with positive definite hessian determines a riemannian metric $\mathbf{g}^D$ and a pair $(\nabla^D, \nabla^{D^\dagger})$ of torsion-free affine connections:

$$\begin{aligned}
\mathbf{g}_\phi(u,v) &:= -\partial_{u|\phi}\partial_{v|\omega} D(\phi,\omega)|_{\omega=\phi},\\
\mathbf{g}_\phi((\nabla_u)_\phi v, w) &:= -\partial_{u|\phi}\partial_{v|\phi}\partial_{w|\omega} D(\phi,\omega)|_{\omega=\phi},\\
\mathbf{g}_\phi(v, (\nabla^\dagger_u)_\phi w) &:= -\partial_{u|\omega}\partial_{w|\omega}\partial_{v|\phi} D(\phi,\omega)|_{\omega=\phi},
\end{aligned}$$

  which satisfy the characteristic equation of the Norden['37]–Sen['44] geometry,

$$\mathbf{g}^D(u,v) = \mathbf{g}^D(\mathbf{t}_c^{\nabla^D}(u), \mathbf{t}_c^{\nabla^{D^\dagger}}(v)) \;\; \forall u,v \in \mathbf{T}\mathcal{M}(\mathcal{N}).$$

- A riemannian geometry $(\mathcal{M}(\mathcal{N}), \mathbf{g}^D)$ has Levi-Civita connection $\bar{\nabla} = (\nabla^D + \nabla^{D^\dagger})/2$.

Hessian geometries = dually flat Norden–Sen geometries

If $(\mathcal{M}, \mathbf{g}, \nabla, \nabla^\dagger)$ is a Norden–Sen geometry with flat ∇ and $\nabla^\dagger$, then:

1. there exists a unique pair of functions $\Phi : \mathcal{M} \to \mathbb{R}$, $\Phi^{\mathsf{L}} : \mathcal{M} \to \mathbb{R}$ such that $\mathbf{g}$ is their hessian metric,

$$\mathbf{g}(\rho) = \sum_{i,j} \frac{\partial^2 \Phi(\rho(\theta))}{\partial\theta^i \partial\theta^j} \mathrm{d}\theta^i \otimes \mathrm{d}\theta^j,$$

$$\mathbf{g}(\rho) = \sum_{i,j} \frac{\partial^2 \Phi^{\mathsf{L}}(\rho(\eta))}{\partial\eta^i \partial\eta^j} \mathrm{d}\eta^i \otimes \mathrm{d}\eta^j,$$

where: $\{\theta^i\}$ is a coordinate system s.t. $\Gamma^{\nabla}_{ijk}(\rho(\theta)) = 0\ \forall\rho \in \mathcal{M}$,
$\{\eta^i\}$ is a coordinate system s.t. $\Gamma^{\nabla^\dagger}_{ijk}(\rho(\eta)) = 0\ \forall\rho \in \mathcal{M}$,
and Φ^{L} is a Fenchel conjugate of Φ.

2. the Eguchi equations applied to the Brègman divergence

$$D_\Phi(\rho, \sigma) := \Phi(\rho) + \Phi^{\mathsf{L}}(\sigma) - \sum_i \theta^i(\rho)\eta^i(\sigma)$$

yield $(\mathbf{g}, \nabla, \nabla^\dagger)$ above.

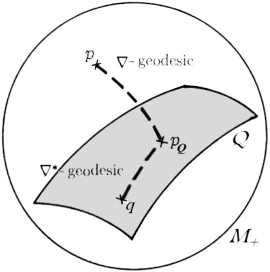
Smooth generalised pythagorean theorem

Let $(\mathcal{M}, \mathbf{g}, \nabla, \nabla^\dagger)$ be a hessian geometry. Then for any $\mathcal{Q} \subseteq \mathcal{M}$ which is:

- $\nabla^\dagger$-autoparallel $:= \nabla^\dagger_u v \in \mathbf{T}\mathcal{Q}\ \forall u, v \in \mathbf{T}\mathcal{Q}$;
- $\nabla^\dagger$-convex $:= \forall \rho_1, \rho_2 \in \mathcal{Q}\ \exists!\ \nabla^\dagger$-geodesics in $\mathcal{Q}$ connecting ρ_1 and ρ_2;

there exists a unique projection

$$\mathcal{M} \ni \rho \mapsto \mathfrak{P}^{D_\Phi}_{\mathcal{Q}}(\rho) := \arg\inf_{\sigma \in \mathcal{Q}} \{D_\Phi(\sigma, \rho)\} \in \mathcal{Q}.$$



- it is equal to a unique projection of ρ onto $\mathcal{Q}$ along a ∇-geodesic that is $\mathbf{g}$-orthogonal at $\mathcal{Q}$.
- it satisfies a generalised pythagorean equation

$$D_\Phi(\omega, \mathfrak{P}^{D_\Phi}_{\mathcal{Q}}(\rho)) + D_\Phi(\mathfrak{P}^{D_\Phi}_{\mathcal{Q}}(\rho), \rho) = D_\Phi(\omega, \rho) \quad \forall(\omega, \rho) \in \mathcal{Q} \times \mathcal{M}.$$

Hence, for Brègman divergences D_Φ the local entropic projections are equivalent with geodesic projections.

Troubles with generality ⇒ localisation via topos models

- Arbitrary/general (effective) state spaces are not dually flat
- Pure states are not faithful and belong only to a boundary of C^∞ manifold
- In such cases as critical points (corresponding to phase transitions) the curvature scalar diverges, corresponding to a singularity of a manifold
- Equations on general form on metric tensor on a manifold represent specific renormalisation constraints [Mitchell'67, Jaynes'85'93, Favretti'07]. This can be interpreted as [RPK'16] "information gravity from renormalisation" (an information geometric 'conceptual analogue' of Sakharov'69+ induced gravity)

Proposed solution [RPK'19]:

- In analogy to a shift from special relativity to general relativity, we can shift to a postulate that hessianity/dual flatness holds infinitesimally (in the same way as Minkowski flatness in GR), or locally [SDG teaches: that's an important difference!]: localisation of 'ideal' inference = 1) locally the default inductive inferences correspond to geodesic falls, 2) infinitesimally, metric has always a hessian form.
- Define hessian algebra as a commutative $\mathbb{R}$-algebra C with a scalar product g and two affine connections ∇ and $\bar{\nabla}$ on a module $\mathbf{T}\mathcal{M}$ of C-derivations s.t. g has positive signature, $\bar{\nabla}g = 0$, $\nabla g = D$, D is completely symmetric third rank tensor on $\mathbf{T}\mathcal{M}$, ∇ and $(\bar{\nabla} - D/2)$ are torsion-free and flat.
- Consider the category HA of hessian algebras with hessian algebra homomorphisms, analogously to category LA of Lorentz algebras, and the topos $\mathbf{Set}^{HA}$.
- Study sheafifications of $\mathbf{Set}^{HA}$, with different subcanonical Grothendieck topologies, encoding different compatibility conditions between local descriptions.

Basic distinction

Two layers of logic: structural vs epistemic (as stressed by Jaynes and Hintikka):

- **deductive** inference:
  - e.g. first order logic
  - premises: formulas that are truth valued (certain)
  - inference: turns certain premises to certain conclusions
- **inductive** inference:
  - e.g. probability theory + Bayes–Laplace rule
  - premises: formulas that are probability valued (plausible)
  - inference: turns plausible premises into most plausible conclusions

Can we extend Lawvere's views on logic with the above distinction?

- "In sheaf theory you do algebra vertically and topology horizontally".
- If: algebra $\leftrightarrow$ logic, and topology $\leftrightarrow$ logic, can the first be deductive, and other inductive?

Three cases:

- GR and toposes:
  1) stage-variation in SDG models as (epistemic) observer's changes of global gauge fixing
  2) Einstein/lorentzian toposes
- QT and categorical brègmannian (nonlinear) geometrostatistics:
  1) global approach: vertically: noncommutative reflexive Banach spaces, horizontally: inductive inference on the spaces of noncommutative integrals,
  2) local approach: dually flat localisation via hessian toposes
- Resource theories and adjointness: epistemic comonads and action monads