Friedrich-Alexander-Universität
Erlangen-Nürnberg

Relative Entropy and the (Quantum) Method of Types

Mathematical Physics Seminar
Universität Regensburg

Ian Koot

June 16 2023

1. The Empirical Distribution and Types

2. The Method of Types

3. A Noncommutative Method of Types?

Probability distributions

Let $X = \{1, 2, \ldots, d\}$ and $P : X \to \mathbb{R}$ a probability distribution. Let

$$P^n(\vec{x}) = P(x_1)P(x_2)\cdots P(x_n).$$

We define the **empirical distribution** associated to $\vec{x}$:

$$ED_n[\vec{x}](i) := \frac{1}{n}\sum_{k=1}^{n} \delta_{i,x_k} = \frac{|\{k : x_k = i\}|}{n}.$$

Intuitively: we expect an outcome whose empirical distribution is 'close to' the true probability distribution. But why exactly? And what does 'close to' mean?

Repeated measurement & ED

The crucial observation is the following:

$$P^n(\vec{x}) = P(x_1)P(x_2)\cdots P(x_n)$$
$$= \prod_{i=1}^{d} P(i)^{nED_n[\vec{x}](i)} = \left(\prod_{i=1}^{d} P(i)^{ED_n[\vec{x}](i)}\right)^n$$

Note:

- $P^n(\vec{x})$ depends only on its empirical distribution.
- For a sequence $\vec{x}_n \in X^n$ with ‘similar’ empirical distributions, the likelihood $P^n(\vec{x}_n)$ will decay exponentially, with a specific rate.

Entropy! And Relative Entropy!

We note:

$$P^n(\vec{x}) = \prod_{i=1}^{d} P(i)^{nED_n[\vec{x}](i)}$$
$$= \prod_{i=1}^{d} \left(\frac{P(i)}{ED_n[\vec{x}](i)}\right)^{nED_n[\vec{x}](i)} ED_n[\vec{x}](i)^{nED_n[\vec{x}](i)}$$

We can write this as follows:

Proposition

Let X be a finite set and $P : X \to \mathbb{R}$ a probability distribution. Then we have

$$P^n(\vec{x}) := \exp\left(-n(S(ED_n[\vec{x}], P) + S(ED_n[\vec{x}]))\right)$$

What happens if $P^n(\vec{x}) = 0$?

Types and Type classes

We group all outcomes by their empirical distribution:

Definition

A probability distribution $P : X \to \mathbb{R}$ is called an n-**type** if there is a $\vec{x} \in X^n$ so that $P = ED_n[\vec{x}]$. It is called a **type** if it is an n-type for some $n \in \mathbb{N}$.

The **type class** associated to the n-type P is the set

$$T_n(P) := \{\vec{x} \in X^n \mid P = ED_n[\vec{x}]\}$$

Note that for any type $P = ED_n[\vec{x}]$ and probability distribution Q we have

$$Q^n(T_n(P)) := \sum_{\vec{y} \in T_n(P)} Q^n(\vec{y}) = |T_n(P)| Q^n(\vec{x})$$

The Method of Types

- Let $i \in X$ be the item for which P is maximal. Then $\vec{x} = (i, i, i, \ldots, i)$ is the most likely result w.r.t. P^n. However, unless $P = \delta_i$, we never see this. Why?
- Answer: *Because the probability of that outcome occurring drops exponentially, and there is only 1 element with that empirical distribution!*
- So more interesting question: which *types* are very likely? And how likely are they? The answer to this question (and applications of this answer) is called the **Method of Types**.

To summarize:

- The **empirical distribution** of an outcome determines its likelihood of occurring;
- The **entropy** and **relative entropy** naturally show up as decay rates for the probability;
- The **method of types** is the method of using the knowledge of which **types** are likely to occur to prove results.
- Specifically, we have

$$P^n(\vec{x}) = \prod_{i}^{d} P(i)^{nED_n[\vec{x}](i)} = e^{-n(S(ED_n[\vec{x}])+S(ED_n[\vec{x}],P))}$$

for

$$S(P) = -\sum_{i=1}^{d} P(i)\ln P(i),$$

$$S(P,Q) = \sum_{i=1}^{d} P(i)\ln P(i) - P(i)\ln Q(i).$$

1. The Empirical Distribution and Types

2. The Method of Types

3. A Noncommutative Method of Types?

Necessary relations

The method of types relies on the following observations:

- $Q^n(T_n(P)) = e^{-nS(P,Q)} P^n(T_n(P))$ for all probability distributions Q and n-types P;
- $P^n(T_n(Q)) \leq P^n(T_n(P))$ for all types n-types P and Q;
- $|ED_n[X^n]| \leq (n+1)^d$;
- $\bigcup_{Q \in ED_n[X^n]} T_n(Q) = X^n$;

Probability asymptotics

Proposition

Let P be an n-type and Q a probability distribution on X. Then

$$\frac{1}{(n+1)^d} e^{-nS(P,Q)} \leq Q^n(T_n(P)) \leq e^{-nS(P,Q)}.$$

Proof.

Note that

$$1 = \sum_{Q \in ED_n[X^n]} P^n(T_n(Q)) \leq (n+1)^d P(T_n(P)).$$

So

$$(n+1)^{-d} \leq P^n(T_n(P)) \leq 1$$

and so the result follows. □

So the decay rate of type P occurring under Q is equal to $S(P,Q)$.

Set Size Asymptotics

Corollary

Let P be an n-type. Then

$$\frac{1}{(n+1)^d} e^{nS(P)} \leq |T_n(P)| \leq e^{nS(P)}.$$

Proof.

If $\vec{x} \in |T_n(P)|$, then $P^n(T_n(P)) = |T_n(P)|P(\vec{x}) = |T_n(P)|e^{-nS(P)}$. So the result follows from

$$(n+1)^{-d} \leq P^n(T_n(P)) \leq 1.$$

□

Note for example the proof of $S(P) \leq \ln d$ for types P, because $|T_n(P)| \leq |X^n| = d^n$ and $S(P) \leq \frac{d \ln(n+1)}{n} + \ln(d) \to \ln(d)$.

Sanov's Theorem

As an immediate application, we can prove **Sanov's theorem**:

Theorem

Let $E \subseteq \mathrm{Pr}(X)$*, and* $P \in \mathrm{Pr}(X)$*. Then*

$$P^n(E) := \sum_{Q \in E \cap ED_n[X^n]} P^n(T_n(Q)) \leq (n+1)^d \sup_{Q \in E} \left(e^{-nS(Q,P)} \right)$$

So: the sets

$$A_{n,\varepsilon}(P) := \{\vec{x} \in X^n \mid S(ED_n[\vec{x}], P) < \varepsilon\}$$

become exponentially likely to occur.

Chernoff-Stein Lemma

We can also strengthen it:

Theorem (Chernoff-Stein Lemma)

Let $P, Q \in \mathrm{Pr}(X)$ *and let* $B_n \subseteq X^n$ *be a sequence of subsets such that* $\lim_{n\to\infty} P(B_n) = 1$*. Then we have*

$$\liminf_{n\to\infty} -\frac{1}{n} \ln Q(B_n) \geq S(P, Q)$$

Furthermore, there is a sequence that achieves this rate (independent of Q*).*

Idea: typical sequences of sets for P (i.e. sequences $B_n \subseteq X^n$ such that $P^n(B_n) \to 1$) must have increasingly large intersections with the entropy typical subsets $A_{n,\varepsilon}(P)$, and those only decay as $e^{-nS(P,Q)}$.

To summarize:

- For P an n-type and Q a probability distribution, we now know that

$$\frac{1}{(n+1)^d}e^{-nS(P,Q)} \leq Q^n(T_n(P)) \leq e^{-nS(P,Q)}$$
$$\frac{1}{(n+1)^d}e^{nS(P)} \leq |T_n(P)| \leq e^{nS(P)}$$

  Compare to $Q_n(\vec{x}) = \exp(-n(S(P) + S(P,Q))$ (for $P = ED_n[\vec{x}]$); a part of the probability is compensated by the size of the type class, a part is not.
- By Sanov's theorem, we see that the sets

$$A_{n,\varepsilon}(P) := \{\vec{x} \in X^n \mid S(ED_n[\vec{x}], P) < \varepsilon\}$$

  will become exponentially likely under P.
- The Chernoff-Stein Lemma tells us that for a P-typical sequence of sets $B_n \subset X^n$ (i.e. such that $\lim_{n\to\infty} P^n(B_n) = 1$), the Q-probability cannot fall off faster than exponentially with rate $S(P,Q)$.

1. The Empirical Distribution and Types

2. The Method of Types

3. A Noncommutative Method of Types?

Noncommutative Probability Theory

We can capture the probability theory of a space of outcomes X with probability distribution P in the following objects:

- The set of functions $L^\infty(X) := \{f : X \to \mathbb{C}\}$.
- The expectation value $\mathbb{E}_P : L^\infty(X) \to \mathbb{C}$ given by $\mathbb{E}_P[f] = \sum_i f(i)P(i)$

[itemsep=8pt] Reconstructing the starting data:
$X \cong \{\delta_x \mid x \in X\} \subseteq L^\infty(X)$ (i.e. set of minimal projections in the algebra) and $P(i) = \mathbb{E}_P[\delta_i]$. Furthermore, sets $A \subseteq X$ correspond to $\chi_A \in L^\infty(X)$, since $P(A) = \mathbb{E}_P[\chi_A]$.

Noncommutative Probability Theory

Drop the assumption of commutativity:

Commutative	Noncommutative
Random variables $L^\infty(X)$	(von Neumann) algebra $\mathcal{A}$
Expectation value $\mathbb{E}_P$	state $\omega \in \mathcal{S}(\mathcal{A})$
Characteristic Funktions χ_A	projections $p \in \mathcal{P}(\mathcal{A})$

If $\mathcal{A}$ has a trace (i.e. a state tr such that $\text{tr}(AB) = \text{tr}(BA)$), then every ω is of the form $\omega(A) = \text{tr}(D_\omega A)$. We also have definitions:

$$S(\omega) = -\text{tr}(D_\omega \ln(D_\omega))$$

$$S(\omega, \psi) = \text{tr}(D_\omega \ln(D_\omega)) - \text{tr}(D_\omega \ln(D_\psi))$$

What doesn't work

Abstractly, we are looking for a map

$$ED_n : \mathcal{P}(\mathcal{A}^{\otimes n}) \supset \mathcal{P}_n \to \mathcal{S}(\mathcal{A}).$$

and a set of projections $T_n(\omega)$ for $\omega \in ED_n[\mathcal{P}_n]$.
We can use our 'dictionary' to translate properties that the classical concepts satisfy:

- $(ED_n[p])^{\otimes n}(p) = e^{-nS(ED_n[p])}$.
- $\omega^{\otimes n}(p) = (ED_n[p])^{\otimes n}(p)e^{-nS(ED_n[p],\omega)}$.
- $ED_n[p]$ maximizes the expression $\omega \mapsto \omega^{\otimes n}(p)$.
- ...

However, many of these properties cannot be realized, are ambiguous or contradict each other. No generally accepted definition exists.

What does work

However, a notion of relative entropy does exist, and the Chernoff-Stein Lemma *does* hold:

Theorem

Let $\mathcal{A}$ *be finite dimensional, and* $\phi, \psi \in \mathcal{S}(\mathcal{A})$. *Then every sequence of projections* $p_n \in \mathcal{P}(\mathcal{A}^{\otimes n})$ *that satisfies* $\lim_{n\to\infty} \psi^{\otimes n}(p_n) = 1$ *also satisfies*

$$\lim_{n\to\infty} -\frac{1}{n}\ln(\phi(p_n)) \leq S(\psi, \phi).$$

Furthermore, there is a sequence p_n *that achieves this rate (this depends on* ϕ *and* ψ*).*

This was proven in [**Bjelakovic2005**] based on results from [**Hiai1991**].

Example

Consider pure states $\omega_v(A) := \langle v, Av \rangle$ and ω_w. Then $S(\omega_w, \omega_v) = \infty$ (if $v \notin \mathbb{C}w$).

- The projections $p_n = |w^{\otimes n}\rangle \langle w^{\otimes n}|$ look 'typical', but $(\omega_v)^{\otimes n}(p_n) = |\langle v, w \rangle|^{2n}$.
- However, we can also let p_n be the projection onto

$$v^{\perp} \otimes w \otimes \ldots \otimes w + w \otimes v^{\perp} \otimes \ldots \otimes w + \ldots .$$

Then $(\omega_w)^{\otimes} n(p_n) \to 1$, and $(\omega_v)^{\otimes n}(p_n) = 0$.

How to prove the NC C.-S. Lemma

Instead, it turns out we can actually reduce to the commutative case. This is because there exists a commutative subalgebra $\mathcal{D}_l$ such that

$$S(\psi^{\otimes l}, \varphi^{\otimes l}) - S(\psi^{\otimes l}|_{\mathcal{D}_l}, \varphi^{\otimes l}|_{\mathcal{D}_l}) \leq |\mathcal{H}| \ln(l+1).$$

For this, if $D_\varphi = \sum_{i=1}^{d} \lambda_i p_i$, we define $p^{\otimes \vec{x}}$ for $\vec{x} \in \{1, \ldots, d\}^n$, and

$$T_l^\varphi(Q) := \bigvee \{p^{\otimes \vec{x}} \mid \vec{x} \in T_l(Q)\}.$$

Then $\mathcal{D}_l$ is the algebra spanned by the eigenprojections of $T_l^\varphi(Q) D_{\psi^{\otimes l}} T_l^\varphi(Q)$.

Summary

To summarize:

- What does not exist (yet): N.C. Empirical Distributions, Types and Type classes. Problem is many ways of being typical.
- What does exist: N.C. (relative) entropy, and the Chernoff-Stein Lemma holds. It says that ψ-typical sequences don't fall off faster in φ-probability then a rate of $S(\psi, \phi)$.
- The crucial observation is that for large enough tensor powers, the noncommutative relative entropy can be approximated by a commutative relative entropy.

Outlook/Questions

Questions we are interested in, are for example:

- How should we interpret the reduction to the commutative subalgebra?
- Could one use the Chernoff-Stein characterization to give more intuitive proofs of the known properties of relative entropy?
- For infinite dimensional algebras, a notion of relative entropy exists. Does the Chernoff-Stein characterization still hold for that setting?
- If so, can we get a better understanding of for example mutual information and entanglement entropy in the QFT setting?
- Modular theory plays a vital role in the definition of relative entropy in infinite dimensions. Can we see this from such a Chernoff-Stein characterization? Can we maybe even learn more about modular theory from this perspective?

Friedrich-Alexander-Universität
Erlangen-Nürnberg

Relative Entropy and the (Quantum) Method of Types

Mathematical Physics Seminar
Universität Regensburg

Ian Koot

June 16 2023