A Note on Distances between Probabilistic and Quantum distributions

Radboud University Nijmegen

Bart Jacobs
bart@cs.ru.nl
June 14, 2017

Where we are, so far

Distances in probability theory

- Much usage of metrics in probabilistic computation, eg.
  - measuring the behavioural similarity of states in probabilistic transition systems
    (Panangaden, Desharnais, Mislove, Worrell, van Breughel, König ...)
  - evaluating performance and uncertainty of Bayesian network models
- Nice systematic description of metrics via universality in LICS'16 paper of Madare, Panangaden, Plotkin
- **Here:** two loosely connected topics, concrete and abstract:
  - "entwinedness" measure for classical and quantum probability
  - metric versions of "state-and-effect" triangles

Discrete probability distributions

Notation

- Fair coin: $\frac{1}{2}|H\rangle + \frac{1}{2}|T\rangle$
- Fair dice: $\frac{1}{6}|1\rangle + \frac{1}{6}|2\rangle + \frac{1}{6}|3\rangle + \frac{1}{6}|4\rangle + \frac{1}{6}|5\rangle + \frac{1}{6}|6\rangle$

ket notation

- $|-\rangle$ is pure syntactic sugar — stemming from quantum
- more confusing to omit them, as in: $\frac{1}{6}1 + \frac{1}{6}2 + \frac{1}{6}3 + \frac{1}{6}4 + \frac{1}{6}5 + \frac{1}{6}6$

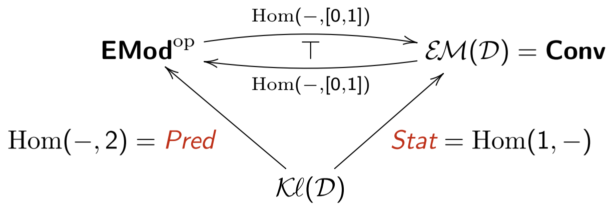
- Write $\mathcal{D}(X)$ for the set of such probability distributions $\sum_i r_i|x_i\rangle$ where $x_i \in X$, $r_i \in [0,1]$ with $\sum_i r_i = 1$
- This $\mathcal{D}$ is a monad on sets; its algebras are convex sets
- Distributions $\omega \in \mathcal{D}(X)$ will often be called states of X

Predicates for probabilistic logic

- A predicate on a set X is a function $p\colon X \to [0,1]$
  - It is called sharp (non-fuzzy) if $p(x) \in \{0,1\}$ for each $x \in X$
- Basic effect module structure on these predicates:
  - true and false, as constant-one and constant-zero
  - orthosupplement $p^{\perp}(x) = 1 - p(x)$
  - partial sum $(p \ovee q)(x) = p(x) + q(x)$ if $p(x) + q(x) \leq 1$ for all x
  - scaling $(r \cdot p)(x) = r \cdot p(x)$, for $r \in [0,1]$
- Each Kleisli map $f\colon X \to \mathcal{D}(Y)$ gives a predicate transformation map $f^{*}\colon [0,1]^{Y} \to [0,1]^{X}$ preserving the effect module structure
- Also sequential and parallel conjunction, via multiplication

State-and-effect triangle

Much is summarised in:



This is a much more general pattern in state- and predicate-transformation semantics of computation — including quantum

Combining states and predicates

Let $\omega \in \mathcal{D}(X)$ be state/distribution, $p \in [0,1]^{X}$ a predicate, both on X.

- **Validity** $\omega \models p$, in $[0,1]$
  - defined as $\sum_{x} \omega(x) \cdot p(x)$
  - also known as expected value of p in state ω
- **Conditioning** $\omega|_{p}$, in $\mathcal{D}(X)$
  - assuming validity $\omega \models p$ is non-zero
  - defined as: $\omega|_{p} = \sum_{x} \frac{\omega(x) \cdot p(x)}{\omega \models p} \big| x \big\rangle$

Validity and conditioning example

- Take $X = \{1,2,3,4,5,6\}$ with state dice $\in \mathcal{D}(X)$
  - recall dice $= \frac{1}{6}|1\rangle + \frac{1}{6}|2\rangle + \frac{1}{6}|3\rangle + \frac{1}{6}|4\rangle + \frac{1}{6}|5\rangle + \frac{1}{6}|6\rangle$
- Take even predicate $E \in [0,1]^{X}$; it's sharp, given by:
  - $E(1) = E(3) = E(5) = 0, \;\; E(2) = E(4) = E(6) = 1$
  - define odd via orthosupplement: $O = E^{\perp}$
- dice $\models E \;=\; \frac{1}{2}$
- $\text{dice}|_{E} \;=\; \frac{1/6}{1/2}|2\rangle + \frac{1/6}{1/2}|4\rangle + \frac{1/6}{1/2}|6\rangle = \frac{1}{3}|2\rangle + \frac{1}{3}|4\rangle + \frac{1}{3}|6\rangle$
- $\text{dice}|_{E} \models O \;=\; 0$

A distance question, asked with Fabio Zanasi (MFCS'17)

- How different are ω and $\omega|_p$?
- This can be formalised as $d(\omega, \omega|_p)$
- This number captures the influence of p on ω
- Used to describe d-separation and blocking of influence in Bayesian networks

Products and marginalisations of states

- For states $\omega_1 \in \mathcal{D}(X_1)$ and $\omega_2 \in \mathcal{D}(X_2)$ we can form the product state $\omega_1 \otimes \omega_2 \in \mathcal{D}(X_1 \times X_2)$ by:

$$\omega_1 \otimes \omega_2 = \sum_{(x_1,x_2)} \big(\omega_1(x_1) \cdot \omega_2(x_2)\big) \big| x_1, x_2 \big\rangle$$

- For a joint state $\sigma \in \mathcal{D}(X_1 \otimes X_2)$ there are marginalisations $M_i(\sigma) \in \mathcal{D}(X_i)$, given by:

$$M_1(\sigma) = \sum_{x_1} \big(\textstyle\sum_{x_2} \sigma(x_1,x_2)\big)\big| x_1 \big\rangle \quad M_2(\sigma) = \sum_{x_2} \big(\textstyle\sum_{x_1} \sigma(x_1,x_2)\big)\big| x_2 \big\rangle$$

- It is to easy that marginalisation after product returns the originals:

$$M_1(\omega_1 \otimes \omega_2) = \omega_1 \qquad M_2(\omega_1 \otimes \omega_2) = \omega_2$$

- But what about the other way around: product of marginals?

Entwinedness (cf. entanglement), and distance

- In general: $\sigma \neq M_1(\sigma) \otimes M_2(\sigma)$
  - "the whole is more than the sum of its parts"
- Example: $\sigma = \frac{1}{2}| a, 1 \rangle + \frac{1}{2}| b, 2 \rangle$, then:
  - $M_1(\sigma) = \frac{1}{2}| a \rangle + \frac{1}{2}| b \rangle$, $M_2(\sigma) = \frac{1}{2}| 1 \rangle + \frac{1}{2}| 2 \rangle$
  - And: $M_1(\sigma) \otimes M_2(\sigma) = \frac{1}{4}| a, 1 \rangle + \frac{1}{4}| a, 2 \rangle + \frac{1}{4}| b, 1 \rangle + \frac{1}{4}| b, 2 \rangle$
- Question: how different are σ and $M_1(\sigma) \otimes M_2(\sigma)$?
  - What is their distance $d\big(\sigma, M_1(\sigma) \otimes M_2(\sigma)\big)$?
  - Can it maximalised?
- How does this compare to the quantum case, where the floklore opinion is that joint states can be "more entangled" than in classical probability?
- The paper contains several experiments
  - findings: the difference is not that big, certainly not in the limit

Broader perspective

Entwinedness of joint states is related to both:

(1) **disintegration** (see also REPAS talk)
- extracting a channel $X \to Y$ from a joint state on X, Y
- (or the other way around)
- deep questions abour correlation and causality
- ongoing joint work with Kenta Cho

(2) **cross-over influence**
- conditioning in one coordinate changes the other coordinate
- in the quantum world, this is Einsteins: "spooky interaction"
- ongoing joint work with Fabio Zanasi (see again MFCS'17)

Where we are, so far

Which distance to use?

- In the **discrete classical** case: total variation distance
  - $\mathbf{tvd}(\omega, \rho) = \frac{1}{2}\sum_x |\omega(x) - \rho(x)|$
  - special (discrete) case of Kantorovic metric, for distributions on metric spaces
- In the **quantum** case: trace distance
  - states are density matrices on Hilbert spaces, or more generally, points of von Neumann algebras
  - $\mathbf{trd}(\omega, \rho) = \frac{1}{2}\mathrm{tr}(|\omega - \rho|) = \frac{1}{2}\mathrm{tr}(\sqrt{(\omega-\rho)^\dagger(\omega-\rho)})$
  - total variation is a special case, for classical (diagonal) states

Both distances have values in the unit interval $[0, 1]$

Both distances have been implemented in EfProb, an embedded language of Python for probabilistic computing (discrete, continuous, quantum)

- See `efprob.cs.ru.nl` and the CALCO tools talk

Experimental results I

- For the 2×2 Bell state qs
  - distance $\mathbf{trd}(qs, M_1(qs) \otimes M_2(qs)) = \frac{3}{4}$
- For a classical 2×2 analogue cs
  - $cs = \frac{1}{2}|0,0\rangle + \frac{1}{2}|1,1\rangle$
  - distance $\mathbf{tvd}(cs, M_1(cs) \otimes M_2(cs)) = \frac{1}{2}$

Experimental results II: n-ary generalisation

- For the $2^n \times 2^n$ n-ary Bell state qs_n
  - distance $\mathbf{trd}(qs_n, M_1(qs_n) \otimes \cdots \otimes M_n(qs_n)) = \frac{2^n - 1}{2^n}$
- For a classical $2^n \times 2^n$ analogue cs_n
  - $cs_n = \frac{1}{2}|0,\ldots,0\rangle + \frac{1}{2}|1,\ldots,1\rangle$
  - distance $\mathbf{tvd}(cs_n, M_1(cs_n) \otimes \cdots \otimes M_n(cs_n)) = \frac{2^{n-1} - 1}{2^{n-1}}$

Conclusion: the classical case legs one step behind the quantum case

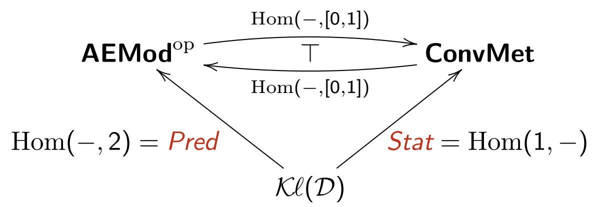
- Aside: the same happens if you use mutual information
- $qs_n \mapsto n \quad cs_n \mapsto n - 1$

Logical reformulation of distances

- In the **classical** case, for Kantorovic distance **kvd**
  - consider distributions ω, ρ on metric space X
  - $\mathbf{kvd}(\omega, \rho) = \bigvee_{\substack{q\colon X \to [0,1] \\ \text{non-expansive}}} \big| \omega \models q - \rho \models q \big|$
  - (alternative formulation uses 'couplings')
- In **quantum** case, for trace distance **trd**
  - for states ω, ρ on Hilbert space $\mathscr{H}$
  - $\mathbf{trd}(\omega, \rho) = \bigvee_{\substack{q\colon \mathscr{H} \to \mathscr{H} \\ 0 \leq q \leq 1}} \big| \omega \models q - \rho \models q \big|$
  - The same definition transfers to von Neumann algebras

Metric state-and-effect triangle

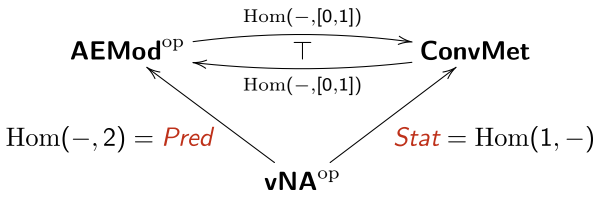
The earlier triangle specialises to:



where:

- **AEMod** is the category of Archimedean effect modules
  - a metric can be defined on them; it's $\bigvee$-metric for predicates
  - so that all effect module maps are automatically non-expansive
- **ConvMet** is the category of convex metric spaces
  - algebra $\mathcal{D}(X) \to X$ must be non-expansive
  - homomorphisms are both affine and non-expansive

Quantum version



Where we are, so far

Final remarks

- Metrics have been used to compare classical and quantum probability, wrt. entwinedness
- Categorically, they fit in the same triangle pictures
- Open question: can we turn the adjunction $\mathbf{AEMod}^{\mathrm{op}} \leftrightarrows \mathbf{ConvMet}$ into a Kadison style duality, by additionally requiring completeness of metric spaces?