Distances in probability theory

- Much usage of metrics in probabilistic computation, e.g.
  - 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: $\frac{1}{2}\ket{H} + \frac{1}{2}\ket{T}$
- Fair dice: $\frac{1}{6}\ket{1} + \frac{1}{6}\ket{2} + \frac{1}{6}\ket{3} + \frac{1}{6}\ket{4} + \frac{1}{6}\ket{5} + \frac{1}{6}\ket{6}$

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

- Write $\mathcal{D}(X)$ for the set of such probability distributions $\sum_i r_i \ket{x_i}$ 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 $p: X \to [0,1]$.
  - 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 $(p \oplus q)(x) = p(x) + q(x)$ 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 $f: X \to D(Y)$ gives a **predicate transformation** $f^*: [0,1]^Y \to [0,1]^X$ preserving the effect module structure
- Also sequential and parallel conjunction, via multiplication

Combining states and predicates

Let $\omega \in D(X)$ 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 $\omega$
- **Conditioning** $\omega|_p$, in $D(X)$
  - assuming validity $\omega \models p$ is non-zero
  - defined as: $\omega|_p = \sum_x \frac{\omega(x) \cdot p(x)}{\omega \models p} | x \rangle$

State-and-effect triangle

Much is summarised in:

\[
\begin{align*}
\text{EMod} & \quad \Leftrightarrow \quad \text{Conv} \\
\text{Hom}(\cdot, [0,1]) & \quad \Leftrightarrow \quad \text{Stat} = \text{Hom}(1, -) \\
\text{Conv} & \quad \Leftrightarrow \quad \text{Stat} = \text{Hom}(1, -)
\end{align*}
\]

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

Validity and conditioning example

- Take $X = \{1,2,3,4,5,6\}$ with state dice $\in D(X)$
  - recall dice $= \frac{1}{6} \left| 1 \right\rangle + \frac{1}{6} \left| 2 \right\rangle + \frac{1}{6} \left| 3 \right\rangle + \frac{1}{6} \left| 4 \right\rangle + \frac{1}{6} \left| 5 \right\rangle + \frac{1}{6} \left| 6 \right\rangle$
- 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: $O = E^\perp$
- $\text{dice} \models E = \frac{1}{2}$
- $\text{dice}\mid_E = \frac{1}{12} \left| 2 \right\rangle + \frac{1}{12} \left| 4 \right\rangle + \frac{1}{12} \left| 6 \right\rangle = \frac{1}{2} \left| 2 \right\rangle + \frac{1}{4} \left| 4 \right\rangle + \frac{1}{3} \left| 6 \right\rangle$
- $\text{dice}\mid_E \models O = 0$
A distance question, asked with Fabio Zanasi (MFCS’17)

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

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 $\sigma$ and $M_1(\sigma) \otimes M_2(\sigma)$?
  - What is their distance $d(\sigma, M_1(\sigma) \otimes M_2(\sigma))$?
  - Can it maximised?
- How does this compare to the quantum case, where the folklore 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

Products and marginalisations of states

- For states $\omega_1 \in D(X_1)$ and $\omega_2 \in D(X_2)$ we can form the product state $\omega_1 \otimes \omega_2 \in D(X_1 \times X_2)$ by:
  $$\omega_1 \otimes \omega_2 = \sum_{(x_1,x_2)} (\omega_1(x_1) \cdot \omega_2(x_2)) |x_1,x_2\rangle$$
- For a joint state $\sigma \in D(X_1 \otimes X_2)$ there are marginalisations $M_i(\sigma) \in D(X_i)$, given by:
  - $M_1(\sigma) = \sum_{x_2} (\sum_{x_1} \sigma(x_1,x_2)) |x_1\rangle$
  - $M_2(\sigma) = \sum_{x_1} (\sum_{x_2} \sigma(x_1,x_2)) |x_2\rangle$
- It is easy that marginalisation after product returns the originals:
  - $M_1(\omega_1 \otimes \omega_2) = \omega_1$
  - $M_2(\omega_1 \otimes \omega_2) = \omega_2$
- But what about the other way around: product of marginals?

Broader perspective

Entwinedness of joint states is related to both:

1. **Disintegration** (see also REPAS talk)
   - extracting a channel $X \rightarrow Y$ from a joint state on $X$, $Y$
   - (or the other way around)
   - deep questions about 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

Introduction

Distances

Conclusions

Experimental results I

- For the $2 \times 2$ Bell state $qs$
  - distance $\text{trd}(qs, M_1(qs) \otimes M_2(qs)) = \frac{3}{4}$

- For a classical $2 \times 2$ analogue $cs$
  - $cs = \frac{1}{2}|0,0\rangle + \frac{1}{2}|1,1\rangle$
  - distance $\text{tvd}(cs, M_1(cs) \otimes M_2(cs)) = \frac{1}{2}$

Experimental results II: $n$-ary generalisation

- For the $2^n \times 2^n$ $n$-ary Bell state $qs_n$
  - distance $\text{trd}(qs_n, M_1(qs_n) \otimes \cdots \otimes M_n(qs_n)) = \frac{2^n - 1}{2^n}$

- 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 $\text{tvd}(cs_n, M_1(cs_n) \otimes \cdots \otimes M_n(cs_n)) = \frac{2^{n-1} - 1}{2^{n-1}}$

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

Aside: the same happens if you use mutual information

$q_{5n} \mapsto n \quad cs_{5n} \mapsto n - 1$
Logical reformulation of distances

- In the classical case, for Kantorovic distance $\kappa V D$
  - consider distributions $\omega, \rho$ on metric space $X$
  - $\kappa V D(\omega, \rho) = \bigvee_{q: X \to [0,1]} |\omega \models q - \rho \models q|$
    - non-expansive
  - (alternative formulation uses 'couplings')

- In quantum case, for trace distance $\text{trd}$
  - for states $\omega, \rho$ on Hilbert space $\mathcal{H}$
  - $\text{trd}(\omega, \rho) = \bigvee_{q: \mathcal{H} \to \mathcal{H}} |\omega \models q - \rho \models q|$
    - $0 \leq q \leq 1$
  - The same definition transfers to von Neumann algebras

Quantum version

- $\text{AEMod}^{\text{op}} \xrightarrow{\text{Hom}(\cdot, [0,1])} \text{ConvMet}^{\text{op}}$
- $\text{Hom}(\cdot, 2) = \text{Pred}$
- $\text{Stat} = \text{Hom}(1, \cdot)$
- $\text{vNA}^{\text{op}} \xleftarrow{\text{Hom}(\cdot, [0,1])} \text{Stat}$

Metric state-and-effect triangle

The earlier triangle specialises to:

- $\text{AEMod}$ 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
- $\text{ConvMet}$ is the category of convex metric spaces
  - algebra $\mathcal{D}(X) \to X$ must be non-expansive
  - homomorphisms are both affine and non-expansive

Where we are, so far

- Introduction
- Distances
- Conclusions
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 $\text{AEMod}^{\text{op}} \leftrightarrow \text{ConvMet}$ into a Kadison style duality, by additionally requiring completeness of metric spaces?