Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in Theoretical Computer Science 336 (2018) 173–187

Electronic Notes in Theoretical Computer Science

www.elsevier.com/locate/entcs

A Note on Distances between Probabilistic and Quantum distributions

Bart Jacobs[1]

Institute for Computing and Information Sciences (iCIS)
Radboud University Nijmegen
The Netherlands

Abstract

This paper is about distances between (or, metrics on) probability distributions. It considers the total variation and Kantorovich distance for discrete distributions, and also the trace distance for quantum distributions. Both concrete and abstract results are presented, showing similarities between classical (discrete) probability and quantum probability. The concrete results involve distances between joint distributions and the product of their marginals, as measure of correlation. It is shown that the discrete and quantum case are not that different. The abstract results address metric aspects of 'state-and-effect' triangles that capture the essential semantic structures at hand.

Keywords: Probability distributions, total variation and Kantorovic distance, trace distance.

1 Introduction

Metric structures have a long history in program semantics, see the overview book [1]. Metrics occur naturally, for instance on sequences, of inputs, outputs, or states. In complete metric spaces solutions of recursive (suitably contractive) equations exist via Banach's fixed point theorem. The Hausdorff distance on subsets is used to model non-deterministic (possibilistic) computation.

This paper looks at metrics on probability distributions. It stands out by comparing standard distance functions on *classical* discrete probability distributions and on *quantum* distributions. The standard distance function that we consider for discrete probability is the total variation distance, which is a special case of the Kantorovich distance, see *e.g.* [6,3,17,2,16]. The distance that we study on quantum distributions is the trace distance. It is, in a sense, also a generalisation of the total variation distance.

We use both these distances in an experiment. Consider a joint distribution ω, which can be discrete or quantum. In both cases we can form its first and second marginal distributions, abbreviated as $\omega_1 = \mathsf{M}_1(\omega)$ and $\omega_2 = \mathsf{M}_2(\omega)$, where M_i is the marginalisation operation. We can put these two marginals together in a product distribution, written as $\omega_1 \otimes \omega_2$.

We ask ourselves the simple question: how much does the joint distribution ω differ from the product of its marginals $\omega_1 \otimes \omega_2$? We measure this difference by

[1] Email: bart@cs.ru.nl

https://doi.org/10.1016/j.entcs.2018.03.022
1571-0661/

taking the distance $d(\omega,\, \omega_1 \otimes \omega_2)$. We call this the *entwinedness* measure. It can be seen as a measure of the correlation (or entanglement) that exists within the joint state ω.

Below in Subsection 2.3 we calculate this distance for a maximally entwined discrete distribution and also for a maximally entwined/entangled quantum distribution. In the quantum case we take the Bell state. Without already spoiling the story too much: there are notable differences between classical and quantum distributions, but the differences become less and less when we move to n-ary products.

These distance calculations quickly become quite complicated, certainly for quantum distributions. The new *EfProb* [2] library [4] does the work for us, and allows us to easily calculate distances for larger distributions. This was of great help for discovering the patterns that are described in Lemmas 2.4 and 2.6.

These concrete distance calculations are a bit of a curiosity. This paper also contains more systematic results about discrete and quantum distances, in Section 3 and 4, where the logical reformulation of these distances in terms of validity $\models$ plays a crucial role. There, metric versions are described of the state-and-effect triangles that emerge in the effectus-theoretic [8,5] description of state and predicate transformer semantics for probability.

2 Distance basics

This section recalls the definitions of *total variation* distance, for discrete probability distributions, and the more general *trace* distance for quantum probability distributions. These distances are used here to describe the distance between a joint distribution and the product of its marginals. This distance can be seen as a measure for the level of 'entwinedness' (or entanglement). It is shown that the classical and quantum cases are rather similar, certainly in the limit.

2.1 Total variation distance

A finite discrete probability distribution on a set X is given by 'probability mass' function $\omega\colon X \to [0,1]$ with finite support and $\sum_x \omega(x) = 1$. This support $\mathrm{supp}(\omega) \subseteq X$ is the set $\{x \in X \mid \omega(x) \neq 0\}$. We often simply say 'distribution' instead of 'finite discrete probability distribution'. Sometimes such a distribution is also called a 'state'. We write $\mathcal{D}(X)$ for the set of distributions on a set X. The mapping $X \mapsto \mathcal{D}(X)$ is a well-known monad see *e.g.* [7,10,11].

The 'ket' notation $|\,-\,\rangle$ is useful to describe specific distributions. For instance, on a set $X = \{a, b, c\}$ we may write a distribution as $\omega = \frac{1}{2}|\,a\,\rangle + \frac{1}{8}|\,b\,\rangle + \frac{3}{8}|\,c\,\rangle$. This corresponds to the probability mass function $\omega\colon X \to [0,1]$ given by $\omega(a) = \frac{1}{2}$, $\omega(b) = \frac{1}{8}$ and $\omega(c) = \frac{3}{8}$.

Definition 2.1 *Let $\omega_1, \omega_2 \in \mathcal{D}(X)$ be two distributions on the same set X. Their*

[2] See efprob.cs.ru.nl

total variation distance $\mathsf{tvd}(\omega_1, \omega_2)$ *is the positive real number defined as:*

$$\mathsf{tvd}(\omega_1, \omega_2) = \frac{1}{2} \sum_{x \in X} \big| \, \omega_1(x) - \omega_2(x) \, \big|. \tag{1}$$

The historical origin of this definition is not precisely clear. It is folklore that the total variation distance is a special case of the 'Kantorovich distance' (also known as 'Wasserstein' or 'earth mover's distance') on distributions on metric spaces, when applied to discrete metric spaces (sets), see Section 3.

We leave it to the reader to verify that tvd is a metric on sets of distributions $\mathcal{D}(X)$, and that its values are in the unit interval $[0, 1]$.

We shall be especially interested in product states $\omega_1 \otimes \omega_2$ and in marginals $\mathsf{M}_1(\sigma), \mathsf{M}_2(\sigma)$ of a joint state σ, defined on a product set like $X_1 \times \cdots \times X_n$. We recall the standard definitions.

For states $\omega_1 \in \mathcal{D}(X_1), \omega_2 \in \mathcal{D}(X_2)$ there is a (joint) product state $\omega_1 \otimes \omega_2 \in \mathcal{D}(X_1 \times X_2)$ given by $(\omega_1 \otimes \omega_2)(x_1, x_2) = \omega_1(x_1) \cdot \omega_2(x_2)$.

In the other direction, for a 'joint' state $\sigma \in \mathcal{D}(X_1 \times X_2)$ there are first and second marginal states $\mathsf{M}_1(\sigma) \in \mathcal{D}(X_1)$ and $\mathsf{M}_2(\sigma) \in \mathcal{D}(X_2)$ given by:

$$\mathsf{M}_1(\sigma)(x_1) = \textstyle\sum_{x_2} \sigma(x_1, x_2) \qquad \mathsf{M}_2(\sigma)(x_2) = \textstyle\sum_{x_1} \sigma(x_1, x_2).$$

We call a joint state σ *non-entwined* when it is the product of its marginals, that is, when $\sigma = \mathsf{M}_1(\sigma) \otimes \mathsf{M}_2(\sigma)$. Every product state $\omega_1 \otimes \omega_2$ is non-entwined, since $\mathsf{M}_i(\omega_1 \otimes \omega_2) = \omega_i$.

The following results are not needed in the sequel, but are worth making explicit.

$$\mathsf{tvd}\big(\omega_1 \otimes \rho, \omega_2 \otimes \rho\big) = \mathsf{tvd}\big(\omega_1, \omega_2\big) \qquad \mathsf{tvd}\big(\mathsf{M}_1(\sigma), \mathsf{M}_1(\tau)\big) \leq \mathsf{tvd}\big(\sigma, \tau\big). \tag{2}$$

2.2 Trace distance

We shall only consider quantum distributions (states) in the finite-dimensional case. For a number $n \in \mathbb{N}$ we write M_n for the set of square $n \times n$ matrices with entries in the complex numbers. A quantum distribution of dimension n is a matrix $\rho \in M_n$ which is positive and has trace equal to one: $\mathrm{tr}(\rho) = 1$, where the trace is the sum of all elements on the diagonal. Such a quantum distribution is often called a (quantum) state. We refer to for instance [18,19,21] for more information.

Definition 2.2 Let $\rho_1, \rho_2 \in M_n$ be two quantum states of the same dimension n. The trace distance $\mathsf{trd}(\rho_1, \rho_2)$ between them is defined as:

$$\mathsf{trd}(\rho_1, \rho_2) = \tfrac{1}{2}\mathrm{tr}\big(\big| \, \rho_1 - \rho_2 \, \big|\big) = \tfrac{1}{2}\mathrm{tr}\Big(\sqrt{(\rho_1 - \rho_2)^\dagger (\rho_1 - \rho_2)}\Big). \tag{3}$$

This definition involves the absolute value $|A|$ of a matrix $A \in M_n$ which is defined as the (matrix) square root of the product $A^\dagger A$, where $(-)^\dagger$ is the conjugate transpose. The square root of a (self-adjoint) matrix B can be computed by first diagonalising the matrix as $B = VDV^\dagger$, where D is a diagonal matrix; then one forms the diagonal matrix $\sqrt{D}$ by taking the square roots of the elements on the diagonal in D; finally the square root of B is $V\sqrt{D}V^\dagger$. Calculating the trace distance by hand is rather unpleasant, but we shall compute it via a tool.

The trace distance is an extension of the total variation distance: given two discrete distributions ω_1, ω_2 on the same set, then the union of their supports $\mathrm{supp}(\omega_1) \cup \mathrm{supp}(\omega_2)$ is a finite set, say with n elements. We can represent ω_1, ω_2 as diagonal matrices $\widehat{\omega_1}, \widehat{\omega_2} \in M_n$. They are states, by construction. Then $\mathsf{trd}(\widehat{\omega_1}, \widehat{\omega_2}) = \mathsf{tvd}(\omega_1, \omega_2)$.

Given two states $\rho_1 \in M_{n_1}$ and $\rho_2 \in M_{n_2}$ one can form the product state $\rho_1 \otimes \rho_2 \in M_{n_1 \cdot n_2}$ via the Kronecker (tensor) product. In the other direction, given a 'joint' state $\tau \in M_{n_1 \cdot n_2}$ one can form the two marginals $\mathsf{M}_1(\tau) \in M_{n_1}$ and $\mathsf{M}_2(\tau) \in M_{n_2}$ via partial traces. We call τ *non-entwined* when it is the tensor product of its marginals.

(The name "entangled" is common in quantum theory, but it has a slightly different meaning than "entwined": a joint state τ is non-entangled when it can be written as *finite* sum of product (sub)states. In the current setting this difference does not matter.)

The facts (2) also hold in the quantum case.

2.3 Calculating entwinedness

It is often claimed that entanglement — or entwinedness, as we shall say here — is a typical quantum phenomenon. But classical (discrete) states can be entwined too. The claim is sometimes re-stated as quantum states can be *more* entangled or *more* correlated than classical states. Our aim in this subsection is to investigated this matter, in terms of (total variation and trace) distance.

The idea is to look at the difference between a joint state and the product of its marginals. We interpret this as a measure of 'entwinedness'. We shall sometimes call this the 'entwinedness' measure.

In order to do the computations we use the *EfProb* library[3] which provides convenient uniform operations for both classical and quantum probability. We only need a small part of this *EfProb* library, namely the part dealing with product states and marginals. The product of two states `s1` and `s2` is written as `s1 @ s2`. This same notation works for discrete and quantum probability. The first and second marginal of a joint state `t` is written via a post-fix operation as `t % [1,0]` and `t % [0,1]`. The selector list [..], also called 'mask', may be of arbitrary length; it describes the components that should be projected/marginalised/traced out via a 0, and the parts that should remain via a 1. We shall also use multi-dimensional states `s1 @...@ sn` and n-ary marginals `t % [0,..,0,1,0,...,0]`.

The Bell state is 'maximally entangled'. Hence it forms an interesting starting point, in order to find out what this difference is between the Bell state and the product of this marginals. You may want to stop here for a moment and think for

[3] Publicly available at efprob.cs.ru.nl

yourself what distance (in the unit interval) you expect.

As a vector in $\mathbb{C}^2 \otimes \mathbb{C}^2$ the Bell state is usually described as $|b\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. In *EfProb* the corresponding density matrix $|b\rangle\langle b| \in M_4$ is called `bell`. The *EfProb*/*Python* code fragments below are hopefully self-explanatory. The part after >>> is typed on the command line, after loading the relevant *EfProb* files.

```
>>> bell
[[ 0.5  0.   0.   0.5]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.5  0.   0.   0.5]]
>>> trdist( bell, (bell % [1,0]) @ (bell % [0,1]) )
0.75
```

This 0.75 is a relatively high distance, which is probably expected.

Let's see what we can do in the discrete case. We take a maximally entwined classical joint state on $\{0, 1\} \times \{0, 1\}$, with the name `cs2`. In the code fragment below this state is printed using ket notation; subsequently its 'entwinedness' measure is produced.

```
>>> cs2
0.5|0,0> + 0|0,1> + 0|1,0> + 0.5|1,1>
>>> tvdist( cs2, (cs2 % [1,0]) @ (cs2 % [0,1]) )
0.5
```

Ha! This classical distance 0.5 is less than the earlier quantum distance 0.75, suggesting indeed that there is classically less correlation / entwinedness / entanglement than in the quantum case.

But let's look a bit further, and consider products of dimension $2 \times 2 \times 2 = 8$. The main candidate now is the 'maximally entangled' GHZ state in M_8. Like the Bell state, it is pre-defined in *EfProb*:

```
>>> ghz
[[ 0.5  0.   0.   0.   0.   0.   0.   0.5]
 [ 0.   0.   0.   0.   0.   0.   0.   0. ]
 [ 0.   0.   0.   0.   0.   0.   0.   0. ]
 [ 0.   0.   0.   0.   0.   0.   0.   0. ]
 [ 0.   0.   0.   0.   0.   0.   0.   0. ]
 [ 0.   0.   0.   0.   0.   0.   0.   0. ]
 [ 0.   0.   0.   0.   0.   0.   0.   0. ]
 [ 0.5  0.   0.   0.   0.   0.   0.   0.5]]
>>> trdist( ghz,
...         (ghz % [1,0,0]) @ (ghz % [0,1,0]) @ (ghz % [0,0,1]) )
0.875
```

We see several interesting things when we compare these Bell and GHZ examples.

- In both cases the density matrices contain zero's `0` everywhere, except for values `0.5` at each of the four corner positions.
- The Bell state of dimension $4 = 2^2$ is compared to a 2-product of its two marginals; the GHZ state of dimension $8 = 2^3$ is compared to the 3-product of its three marginals.
- The 'entwinedness' measure in the Bell case is $\frac{3}{4}$, whereas in the GHZ case it is higher, namely $\frac{7}{8}$.

We can generalise this to n-products, and conjecture that the entwinedness measure is then $\frac{2^n-1}{2^n}$. This is confirmed by a few more of these distance calculations in *EfProb*, for $n = 4, 5, 6, \ldots$, but after $n = 10$ rounding errors start playing a role, starting with differences in the order of 10^{-3}. Hence it is time to turn to a mathematical description.

Definition 2.3 Write $\mathsf{qs}_n \in M_{2^n}$ for the matrix/state consisting of only 0's, except for its four outer corners, which have entries $\frac{1}{2}$.

Thus the Bell state is qs_2 and the GHZ state is qs_3. It is easy to define qs_n as a function in *Python*, and to obtain a list of its n-marginals of dimension 2. The pattern in the next result emerged via experiments.

Lemma 2.4 *Consider the 'quantum state' qs_n from Definition 2.3, for $n \geq 2$.*

(i) *For each $i \leq n$, the i-th marginal $\mathsf{M}_i(\mathsf{qs}_n)$ is equal to the fair 'quantum coin' state $\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \in M_2$.*

(ii) *The product state $\mathsf{M}_1(\mathsf{qs}_n) \otimes \cdots \otimes \mathsf{M}_n(\mathsf{qs}_n)$ of these n marginals is the diagonal (uniform) state $1/2^n \cdot I \in M_{2^n}$.*

(iii) *The entwinedness measure $\mathsf{trd}\big(\mathsf{qs}_n, \mathsf{M}_1(\mathsf{qs}_n) \otimes \cdots \otimes \mathsf{M}_n(\mathsf{qs}_n)\big)$ equals $\frac{2^n-1}{2^n}$.*

We thus see that the entwinedness measure goes to the maximum value 1 as n goes to infinity.

Proof. For the last point the matrix $\mathsf{qs}_n - \mathsf{M}_1(\mathsf{qs}_n) \otimes \cdots \otimes \mathsf{M}_n(\mathsf{qs}_n)$ is described on the left below, and its absolute (matrix) value on the right.

$$\begin{pmatrix} 1/2 - 1/2^n & 0 & \cdots & 0 & 1/2 \\ 0 & -1/2^n & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & -1/2^n & 0 \\ 1/2 & 0 & \cdots & 0 & 1/2 - 1/2^n \end{pmatrix} \qquad \begin{pmatrix} 1/2 & 0 & \cdots & 0 & 1/2 - 1/2^n \\ 0 & 1/2^n & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1/2^n & 0 \\ 1/2 - 1/2^n & 0 & \cdots & 0 & 1/2 \end{pmatrix}$$

We can now compute the entwinedness measure as half of the trace of the matrix on the right:

$$\tfrac{1}{2}\big(\tfrac{1}{2} + (2^n - 2) \cdot \tfrac{1}{2^n} + \tfrac{1}{2}\big) = \tfrac{1}{2} + \tfrac{2^{n-1}-1}{2^n} = \tfrac{2^{n-1}}{2^n} + \tfrac{2^{n-1}-1}{2^n} = \tfrac{2^n-1}{2^n}. \qquad \square$$

We return to discrete probability. We have seen that the entwinedness measure for the state `cs2` is $\frac{1}{2}$. What happens to this distance as we go to 3-products, 4-products, *etc.*

It is easy to define a 'classical state' function `cs(n)` in *Python* which returns a discrete state on $\{0,1\}^n$ of length 2^n, with $\frac{1}{2}$ at the outside, as in:

```
>>> cs(3)
0.5|0,0,0> + 0|0,0,1> + 0|0,1,0> + 0|0,1,1> + 0|1,0,0> + 0|1,0,1>
   + 0|1,1,0> + 0.5|1,1,1>
>>> tvdist( cs(3),
...          (cs(3) % [1,0,0]) @ (cs(3) % [0,1,0]) @ (cs(3) % [0,0,1]) )
0.75
```

For $n = 4$ the distance is $\frac{7}{8}$, for $n = 5$ it's $\frac{15}{16}$. Hence we expect the pattern to be $\frac{2^{n-1}-1}{2^{n-1}}$.

Definition 2.5 Let $n \geq 1$, and $(b_i)_{i<2^n}$ be the n-length bit words representing the numbers $0, 1, 2, \ldots, 2^n - 1$. Write cs_n for the 'classical state' on the set $\{b_i \mid i < 2^n\}$ with probability $\frac{1}{2}$ for $b_0 = 00\cdots00$ and also for $b_{2^n-1} = 11\cdots11$, and probability 0 everywhere else.

Lemma 2.6 *Consider the classical state* cs_n *for* $n \geq 2$.

(i) *Each of the* n *marginals* $\mathsf{M}_i(\mathsf{cs}_n)$ *is the fair coin* $\frac{1}{2}|0\rangle + \frac{1}{2}|1\rangle$.

(ii) *The product state* $\mathsf{M}_1(\mathsf{cs}_n) \otimes \cdots \otimes \mathsf{M}_n(\mathsf{cs}_n)$ *of these* n *marginals is the uniform distribution on the set* $\{b_i \mid i < 2^n\}$ *with probability* $\frac{1}{2^n}$ *for each bitstring* b_i.

(iii) *The entwinedness measure* $\mathsf{tvd}\big(\mathsf{cs}_n, \mathsf{M}_1(\mathsf{cs}_n) \otimes \cdots \otimes \mathsf{M}_n(\mathsf{cs}_n)\big)$ *equals* $\frac{2^{n-1}-1}{2^{n-1}}$.

We thus see that the classical entwinedness measure also goes asymptotically to 1 as n goes to infinity, but lags one step behind the quantum distance.

Proof. Again we concentrate on the last point. The distance is the sum:

$$
\begin{aligned}
\tfrac{1}{2}\big((\tfrac{1}{2} - \tfrac{1}{2^n}) + (2^n - 2) \cdot \tfrac{1}{2^n} + (\tfrac{1}{2} - \tfrac{1}{2^n})\big) &= \tfrac{1}{2} + \tfrac{1}{2} \cdot (2^n - 4) \cdot \tfrac{1}{2^n} \\
&= \tfrac{2^{n-2}}{2^{n-1}} + \tfrac{2^{n-2}-1)}{2^{n-1}} \\
&= \tfrac{2^{n-1}-1}{2^{n-1}}.
\end{aligned}
$$

□

Our conclusion is that maximally entwined quantum states qs_n do not differ dramatically from the maximally entwined classical (discrete probabilistic) states cs_n when it comes to entwinedness measure — certainly when n goes to infinity.

Remark 2.7 We have investigated the distance $d(\omega, \omega_1 \otimes \cdots \otimes \omega_n)$ between a state ω and the product of its two marginals $\omega_1, \ldots, \omega_n$. In information theory one usually looks at *mutual information*, that is, at the Kullback-Leibler divergence of the product of the marginals, from the joint distribution. Concretely, it gives the expectation of the logarithmic difference:

$$\begin{aligned} D_{KL}(\omega \| \omega_1 \otimes \omega_2) &= \sum\nolimits_{x,y} \omega(x,y) \cdot \log\Big(\frac{\omega(x,y)}{\omega_1(x) \cdot \omega_2(y)}\Big) \\ &= H(\omega_1) + H(\omega_2) - H(\omega). \end{aligned}$$

The latter formulation uses Shannon entropy H, see [18]. It has been implemented in *EfProb*, also for n-ary states. The mutual information value of the quantum states states $\mathsf{qs}(i)$ is i and for classical states $\mathsf{cs}(i)$ it is $i-1$. Hence again we see that the classical states lag one step behind the quantum ones.

Nevertheless, we are not aware of a formal relationship between the entwinedness measure that we use here and mutual information.

3 Kantorovich distance

The example calculations in the previous section are very concrete applications of the total variation and trace distances. In the remaining sections of this paper we will look at these distances from a more general perspective. This section concentrates on the total variation distance, and the next one on the trace distance.

As already mentioned in the previous section, the total variation distance is a special case of the Kantorovich distance on metric spaces. In order to describe the latter, we need some context.

A metric d on a set X is called 1-bounded if it takes values in the unit interval $[0,1]$, that is, if it has type $d\colon X \times X \to [0,1]$. We write **Met** for the category with such 1-bounded metric spaces as objects, and with non-expansive functions f between them, satisfying $d(f(x), f(y)) \leq d(x,y)$. This category **Met** is complete and cocomplete, and monoidal closed. The metric on a cartesian product $\times$ is given by joins, and on a tensor product $\otimes$ by truncated sums.

From now on we assume that all metric spaces in this paper are 1-bounded. For example, each set carries a discrete metric, where points x, y have distance 0 if they are equal, and 1 otherwise.

For a metric space X and two functions $f, g\colon A \to X$ from some set A to X there is the *supremum* distance given by:

$$\mathsf{spd}(f,g) = \bigvee_{a \in A} d\big(f(a), g(a)\big). \tag{4}$$

A *predicate* on a metric space X is a non-expansive function $p\colon X \to [0,1]$. These predicates carry the above supremum distance spd. For a discrete probability distribution ω on X we write $\omega \models p$ for the validity (or expected value) of p in ω. It is defined as the (finite) sum $\sum_x \omega(x) \cdot p(x)$.

Definition 3.1 Let ω_1, ω_2 be two discrete distributions on (the underlying set of) a metric space X. The *Kantorovich* distance between them is defined as:

$$\mathsf{kvd}(\omega_1, \omega_2) = \bigvee_{p \in \mathbf{Met}(X,[0,1])} \Big| \omega_1 \models p - \omega_2 \models p \Big|. \tag{5}$$

This makes $\mathcal{D}(X)$ a (1-bounded) metric space.

The Kantorovich-Wasserstein duality Theorem gives an equivalent description of this distance in terms of joint states and 'couplings', see [15,20] for details. Here we are interested in relating the Kantorovich distance to the monad structure of distributions. For this we really need to move from the total variation distance tvd for distributions on sets to the Kantorovich distance kvd for distributions on metric spaces, namely when we consider distributions of distributions $\mathcal{D}(\mathcal{D}(X))$. Even if X is just a set, $\mathcal{D}(X)$ is a metric space, so we need to take this metric structure into account when we form $\mathcal{D}(\mathcal{D}(X))$.

The following result is standard and is included without proof, but see [6,20] for more information. It shows that the Kantorovich distance kvd and total variation distance tvd coincide on discrete spaces, and can be described in terms of 'sharp' predicates $X \to \{0, 1\}$, taking values in $\{0, 1\}$ instead of in $[0, 1]$. The sharp predicates correspond to subsets $U \subseteq X$, via the indicator function $\mathbf{1}_U : X \to \{0, 1\}$.

Proposition 3.2 *Let X be an arbitrary set, considered as discrete metric space. Then for distributions $\omega_1, \omega_2 \in \mathcal{D}(X)$ one has the following series of equalities.*

$$\begin{aligned}
\mathsf{kvd}\big(\omega_1, \omega_2\big) &\overset{(5)}{=} \bigvee_{p \in [0,1]^X} \Big| \omega_1 \models p - \omega_2 \models p \Big| \\
&= \bigvee_{p \in \{0,1\}^X} \Big| \omega_1 \models p - \omega_2 \models p \Big| \\
&= \bigvee_{U \subseteq X} \Big| \omega_1 \models \mathbf{1}_U - \omega_2 \models \mathbf{1}_U \Big| \\
&= \tfrac{1}{2} \sum_{x \in X} \big| \omega_1(x) - \omega_2(x) \big| \overset{(1)}{=} \mathsf{tvd}\big(\omega_1, \omega_2\big).
\end{aligned}$$

For a Kleisli map $f : X \to \mathcal{D}(Y)$ there are two associated 'transformation' functions, namely state transformation $f_* : \mathcal{D}(X) \to \mathcal{D}(Y)$ and predicate transformation $f^* : [0, 1]^Y \to [0, 1]^X$. State transformation (aka. Kleisli extension) is defined as $f_*(\omega)(y) = \sum_x f(x)(y) \cdot \omega(x)$, and predicate transformation as $f^*(q)(x) = \sum_y f(x)(y) \cdot q(y)$. They satisfy the fundamental validity transformation equality: $f_*(\omega) \models q = \omega \models f^*(q)$.

Lemma 3.3 *Let X, Y be metric spaces.*

(i) *The unit function $\eta : X \to \mathcal{D}(X)$ given by $\eta(x) = 1|x\rangle$ is non-expansive.*

(ii) *For each non-expansive function $f : X \to \mathcal{D}(Y)$ the corresponding state transformer $f_* : \mathcal{D}(X) \to \mathcal{D}(Y)$ is non-expansive. As special cases, the multiplication map $\mu = (\mathrm{id})_* : \mathcal{D}(\mathcal{D}(X)) \to \mathcal{D}(X)$ is non-expansive, and validity $(-) \models p = p_* : \mathcal{D}(X) \to \mathcal{D}(2) = [0, 1]$ in its first argument too.*

(iii) *If* $f\colon X \to \mathcal{D}(Y)$ *and* $q\colon Y \to [0,1]$ *are non-expansive, then so is* $f^*(q)\colon X \to [0,1]$*. Moreover, the function* $f^*\colon \mathbf{Met}(Y,[0,1]) \to \mathbf{Met}(X,[0,1])$ *is itself non-expansive, wrt. the supremum distance. Hence validity* $\omega \models (-) = \omega^*\colon \mathbf{Met}(X,[0,1]) \to \mathbf{Met}(1,[0,1]) = [0,1]$ *is non-expansive in its second argument too.*

(iv) *Taking convex combinations of distributions satisfies: for* $r+s=1$*,*

$$\mathsf{kvd}\big(r\cdot\sigma_1 + s\cdot\sigma_2,\ r\cdot\tau_1 + s\cdot\tau_2\big) \leq r\cdot\mathsf{kvd}(\sigma_1,\tau_1) + s\cdot\mathsf{kvd}(\sigma_2,\tau_2).$$

We conclude that $\mathcal{D}$ *lifts from a monad on the category* **Sets** *to the category* **Met** *as described in:*

$$\begin{tikzcd} \mathbf{Met} \arrow[r, "\mathcal{D}"] \arrow[d] & \mathbf{Met} \arrow[d] \\ \mathbf{Sets} \arrow[r, "\mathcal{D}"] & \mathbf{Sets} \end{tikzcd} \tag{6}$$

The lifting (6) can be seen as a finite version of a similar lifting result for the 'Kantorovich' functor $\mathcal{K}$ in [3]. This $\mathcal{K}(X)$ captures the tight Borel probability measures on a metric space X. The above lifting (6) is a special case of the generic lifting of functors on sets to functors on metric spaces described in [2] (see esp. Example 3.3).

Proof. We do the first and the last point and leave the others to the reader. The crucial point that we use to show that the unit map $\eta\colon X \to \mathcal{D}(X)$ is non-expansive is: $\eta(x) \models p = p(x)$. Hence we are done because the join in (5) is over non-expansive functions p in:

$$\begin{aligned} \mathsf{kvd}(\eta(x_1),\eta(x_2)) \;=\; \bigvee_p \Big|\,\eta(x_1)\models p - \eta(x_2)\models p\,\Big| &= \bigvee_p \Big|\,p(x_1) - p(x_2)\,\Big| \\ &\leq \bigvee_p d(x_1,x_2) \\ &= d(x_1,x_2). \end{aligned}$$

For the last point we first notice that for $\Omega \in \mathcal{D}^2(X)$ and $p\colon X \to [0,1]$,

$$\begin{aligned} \mu(\Omega)\models p \;=\; \textstyle\sum_x \mu(\Omega)(x)\cdot p(x) &= \textstyle\sum_x \big(\sum_\omega \Omega(\omega)\cdot\omega(x)\big)\cdot p(x) \\ &= \textstyle\sum_\omega \Omega(\omega)\cdot\big(\sum_x \omega(x)\cdot p(x)\big) \\ &= \textstyle\sum_\omega \Omega(\omega)\cdot\big(\omega\models p\big) \\ &= \Omega \models \big((-)\models p\big), \end{aligned}$$

where $(-)\models p\colon \mathcal{D}(X) \to [0,1]$ is used as (non-expansive) predicate on $\mathcal{D}(X)$. Hence for $r, s \in [0,1]$ with $r + s = 1$,

$$\begin{aligned}
&\mathsf{kvd}\big(r\cdot\sigma_1 + s\cdot\sigma_2,\ r\cdot\tau_1 + s\cdot\tau_2\big) \\
&= \mathsf{kvd}\big(\mu(r|\sigma_1\rangle + s|\sigma_2\rangle),\ \mu(r|\tau_1\rangle + s|\tau_2\rangle)\big) \\
&= \bigvee\nolimits_p \big|\,\mu(r|\sigma_1\rangle + s|\sigma_2\rangle) \models p - \mu(r|\tau_1\rangle + s|\tau_2\rangle) \models p\,\big| \\
&= \bigvee\nolimits_p \big|\,r|\sigma_1\rangle + s|\sigma_2\rangle \models \big((-) \models p\big) - r|\tau_1\rangle + s|\tau_2\rangle \models \big((-) \models p\big)\,\big| \\
&= \bigvee\nolimits_p \big|\,r\cdot(\sigma_1 \models p) + s\cdot(\sigma_2 \models p) - r\cdot(\tau_1 \models p) + s\cdot(\tau_2 \models p)\,\big| \\
&\leq \bigvee\nolimits_p r\cdot\big|\,\sigma_1 \models p - \tau_1 \models p\,\big| + \bigvee\nolimits_p s\cdot\big|\,\sigma_2 \models p - \tau_2 \models p\,\big| \\
&= r\cdot\mathsf{kvd}(\sigma_1,\tau_1) + s\cdot\mathsf{kvd}(\sigma_2,\tau_2).
\end{aligned}$$

□

In a probabilistic and quantum setting predicates carry the structure of an *effect module*, see [9,8,5]. Writing **EMod** for the category of effect modules, and **Conv** $= \mathcal{EM}(\mathcal{D})$ for the category of convex sets — as Eilenberg-Moore algebras of the distribution monad $\mathcal{D}$ on **Sets** — we have the standard 'state-and-effect' triangle for discrete probability as on the left below, where the adjunction at the top is obtained by 'homming into $[0,1]$'.

$$\begin{tikzcd}
\mathbf{EMod}^{\mathrm{op}} \arrow[rr, bend left, ""] \arrow[rr, phantom, "\top"] & & \mathbf{Conv} \arrow[ll, bend left, ""] \\
& \mathcal{K}\ell(\mathcal{D}) \arrow[ul, "\mathrm{Hom}(-,2)=\mathrm{Pred}"] \arrow[ur, "\mathrm{Stat}=\mathrm{Hom}(1,-)"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbf{AEMod}^{\mathrm{op}} \arrow[rr, bend left, ""] \arrow[rr, phantom, "\top"] & & \mathbf{ConvMet} \arrow[ll, bend left, ""] \\
& \mathcal{K}\ell(\mathcal{D}) \arrow[ul, "\mathrm{Pred}"] \arrow[ur, "\mathrm{Stat}"'] &
\end{tikzcd}
\tag{7}$$

These state and effect triangles provide a systematic pattern where computations are maps f in the basis category, which give rise to forward state transformers $\mathrm{Stat}(f) = f_*$ in **Conv** and backward predicate transformers $\mathrm{Pred}(f) = f^*$ in **EMod**. The adjunction at the top gives the standard dual adjoint relationship between algebraic logics and spaces, see [12] for more information.

Our next aim is to prove that the triangle on the left restricts to the triangle on the right. It involves two subcategories **AEMod** $\hookrightarrow$ **EMod** and **Conv** $\hookleftarrow$ **ConvMet** $\hookrightarrow$ **Met**.

- The category **AEMod** of *Archimedean* effect modules is defined in [13,14]. The precise definition of the Archimedean property in effect modules is a bit subtle: $x \leq y$ follows if $\frac{1}{2}x \leq \frac{1}{2}y \ovee \frac{r}{2}1$ holds for all $r \in (0,1]$. But it leads to a some neat results like:
  - the full subcategory **AEMod** of Archimedean effect modules is equivalent to the category of order unit spaces;
  - Archimedean effect modules carry a (1-bounded) metric, and all maps of effect modules are automatically non-expansive. This gives a functor **AEMod** $\to$ **Met**.

  The metric induced on Archimedean effect modules of fuzzy predicates $[0,1]^X$ is the supremum metric (4).
- The category **ConvMet** contains *convex metric spaces*, consisting of:

(i) a convex set X, that is, a set X with an Eilenberg-Moore algebra $\alpha\colon \mathcal{D}(X) \to X$ of the distribution monad $\mathcal{D}$ on **Sets**;

(ii) a metric $d_X\colon X \times X \to [0,1]$ on X;

(iii) a connection between the convex and the metric structure, via the requirement that the algebra map $\alpha\colon \mathcal{D}(X) \to X$ is non-expansive: $d_X(\alpha(\omega_1), \alpha(\omega_2)) \leq \mathsf{kvd}(\omega_1, \omega_2)$, for all distributions $\omega_1, \omega_2 \in \mathcal{D}(X)$.

The maps in **ConvMet** are both affine and non-expansive. Thus, **ConvMet** is the Eilenberg-Moore category of the lifted monad $\mathcal{D}\colon \mathbf{Met} \to \mathbf{Met}$ in (6).

Example 3.4 The unit interval $[0,1]$ is a convex metric spaces, via its standard (Euclidean) metric, and its standard convex structure, given by the algebra map $\alpha\colon \mathcal{D}([0,1]) \to [0,1]$ defined by the 'expected value' operation:

$$\alpha(\omega) = \textstyle\sum_{x\in\mathbb{R}} \omega(x)\cdot x \qquad \text{that is} \qquad \alpha\big(\sum_i r_i|x_i\rangle\big) = \sum_i r_i \cdot x_i.$$

The identity map $\mathrm{id}\colon [0,1] \to [0,1]$ is a predicate on $[0,1]$ that satisfies:

$$\omega \models \mathrm{id} = \textstyle\sum_x \omega(x)\cdot \mathrm{id}(x) = \sum_x \omega(x)\cdot x = \alpha(\omega).$$

This allows us to show that α is non-expansive:

$$\begin{aligned} \big|\alpha(\omega_1) - \alpha(\omega_2)\big| &= \big|\,\omega_1 \models \mathrm{id} - \omega_2 \models \mathrm{id}\,\big| \\ &\leq \textstyle\bigvee_p \big|\,\omega_1 \models p - \omega_2 \models p\,\big| = \mathsf{kvd}(\omega_1, \omega_2). \end{aligned}$$

Theorem 3.5 *The state-and-effect triangle on the left in* (7) *restricts to the 'metric' one on the right.*

Proof. The 'states' (comparison) functor $\mathrm{Stat}\colon \mathcal{K}\ell(\mathcal{D}) \to \mathcal{EM}(\mathcal{D}) = \mathbf{Conv}$ in (7) restricts to **ConvMet** by Lemma 3.3, since free algebras $\mu\colon \mathcal{D}(\mathcal{D}(X)) \to \mathcal{D}(X)$ are non-expansive (and affine); moreover, state transformers f_* are both non-expansive and affine. The 'predicate' functor $\mathrm{Pred} = [0,1]^{(-)}\colon \mathcal{K}\ell(\mathcal{D}) \to \mathbf{EMod}^{\mathrm{op}}$ restricts to $\mathbf{AEMod}^{\mathrm{op}}$ since sets of predicates $[0,1]^X$ are Archimedean, as remarked above, following [13,14].

We have to show that the adjunction in (7) restricts appropriately. For an effect module E, the homset $\mathrm{Hom}(E,[0,1])$ carries a convex structure that is given by the map

$$\mathcal{D}\big(\mathrm{Hom}(E,[0,1])\big) \xrightarrow{\ \alpha\ } \mathrm{Hom}(E,[0,1]) \quad \text{with} \quad \alpha(\omega)(e) = \textstyle\sum_h \omega(h)\cdot h(e),$$

where h ranges over $\mathrm{Hom}(E,[0,1])$. Notice that each element $e \in E$ gives rise to a predicate $\mathrm{ev}_e\colon \mathrm{Hom}(E,[0,1]) \to [0,1]$ via $\mathrm{ev}_e(h) = h(e)$. It satisfies for $\omega \in \mathcal{D}\big(\mathrm{Hom}(E,[0,1])\big)$,

$$\omega \models \mathrm{ev}_e = \textstyle\sum_h \omega(h)\cdot \mathrm{ev}_e(h) = \sum_h \omega(h)\cdot h(e) = \alpha(\omega)(e).$$

Now we can show that the algebra map α on $\mathrm{Hom}(E, [0,1])$ is non-expansive:

$$\begin{aligned} d\big(\alpha(\omega_1), \alpha(\omega_2)\big) \stackrel{(4)}{=} \bigvee_e \big| \alpha(\omega_1)(e) - \alpha(\omega_2)(e) \big| &= \bigvee_e \big| \omega_1 \models \mathrm{ev}_e - \omega_2 \models \mathrm{ev}_e \big| \\ &\leq \bigvee_p \big| \omega_1 \models p - \omega_2 \models p \big| \\ &= \mathsf{kvd}(\omega_1, \omega_2). \end{aligned}$$

Each map $f\colon E \to D$ in **EMod** gives an affine map $(-) \circ f\colon \mathrm{Hom}(D, [0,1]) \to \mathrm{Hom}(E, [0,1])$ in **Conv**; it is easy to show that it is also non-expansive.

In the other direction we have to prove that for each convex metric set X the set $\mathrm{Hom}(X, [0,1])$ of affine non-expansive maps is Archimedean. This follows from the fact that the set of functions $[0,1]^X$ is Archimedean. □

4 Trace distance

This section describes, in analogy with the previous one, some basic properties of the trace distance on quantum states. Abstractly, a *state* of a C^* or W^* (von Neumann) algebra A is a completely positive map $\rho\colon A \to \mathbb{C}$. A *predicate* (also called 'effect') of A is an element $e \in A$ with $0 \leq e \leq 1$. We write $[0,1]_A \subseteq A$ for the subset of predicates. The *validity* $\rho \models e$ is the probability $\rho(a) \in [0,1]$. A predicate is called *sharp* if $e \cdot e = e$, that is, if e is a projection.

We rely heavily on the following standard result, see *e.g.* [18, §9.2]. It is the quantum analogue of Proposition 3.2.

Proposition 4.1 *Let $A = \mathcal{B}(\mathscr{H})$ be a von Neumann algebra of operators on a finite-dimensional Hilbert space $\mathscr{H}$, with states $\rho_1, \rho_2\colon A \to \mathbb{C}$. Then:*

$$\begin{aligned} \mathsf{trd}(\rho_1, \rho_2) &= \bigvee_{e \in [0,1]_A} \big| \rho_1 \models e - \rho_2 \models e \big| \\ &= \bigvee_{s \in [0,1]_A \ sharp} \big| \rho_1 \models s - \rho_2 \models s \big|. \end{aligned} \tag{8}$$

In the light of this result we take as distance on the set $\mathrm{Stat}(A) = \mathrm{Hom}(A, \mathbb{C})$ of states of a von Neumann algebra A, the join(s) occurring in (8). We will still call this the *trace* distance trd. It is well-known that states are closed under convex combinations, and thus form a convex set, formally via a function $\alpha\colon \mathcal{D}(\mathrm{Stat}(A)) \to \mathrm{Stat}(A)$. With their trace distance they form a metric space too. We will show that the map α is non-expansive.

Lemma 4.2 (i) *Let $e \in [0,1]_A$ be a predicate. The 'evaluate at e' map $\mathrm{ev}_e = (-)(e) = (-) \models e\colon \mathrm{Hom}(A, \mathbb{C}) \to [0,1]$ is both affine and non-expansive.*

(ii) *The convex map $\alpha\colon \mathcal{D}(\mathrm{Stat}(A)) \to \mathrm{Stat}(A)$ is non-expansive.*

(iii) *The 'states' functor $\mathrm{Stat} = \mathrm{Hom}(-, \mathbb{C})\colon \mathbf{vNA}^{\mathrm{op}} \to \mathbf{Conv}$ restricts to $\mathrm{Stat}\colon \mathbf{vNA}^{\mathrm{op}} \to \mathbf{ConvMet}$.*

Proof. (i) It is standard that the map ev_e is affine, so we concentrate on its non-expansiveness: for states ρ_1, ρ_2 we have:

$$\big|\,\mathrm{ev}_e(\rho_1) - \mathrm{ev}_e(\rho_2)\,\big| \;=\; \big|\,\rho_1 \models e - \rho_2 \models e\,\big| \;\leq\; \bigvee\nolimits_{a\in[0,1]_A} \big|\,\rho_1 \models a - \rho_2 \models a\,\big| \stackrel{(8)}{=} \mathsf{trd}(\rho_1, \rho_2)$$

(ii) Suppose we have two formal convex combinations $\Omega = \sum_i r_i|\omega_i\rangle$ and $\Psi = \sum_j s_j|\rho_j\rangle$ in $\mathcal{D}(\mathrm{Stat}(A))$. The map $\alpha\colon \mathcal{D}(\mathrm{Stat}(A)) \to \mathrm{Stat}(A)$ is non-expansive since:

$$\begin{aligned}
\mathsf{trd}\big(\alpha(\Omega), \alpha(\Psi)\big) &\stackrel{(8)}{=} \bigvee\nolimits_e \big|\,(\textstyle\sum_i r_i\cdot\omega_i) \models e - (\sum_j s_j\cdot\rho_j) \models e\,\big| \\
&= \bigvee\nolimits_e \big|\,\textstyle\sum_i r_i\cdot\omega_i(e) - \sum_j s_j\cdot\rho_j(e)\,\big| \\
&= \bigvee\nolimits_e \big|\,\textstyle\sum_i r_i\cdot\mathrm{ev}_e(\omega_i) - \sum_j s_j\cdot\mathrm{ev}_e(\rho_j)\,\big| \\
&= \bigvee\nolimits_e \big|\,\Omega \models \mathrm{ev}_e - \Psi \models \mathrm{ev}_e\,\big| \\
&\stackrel{(i)}{\leq} \bigvee\nolimits_{p\in\mathbf{Met}(\mathrm{Stat}(A),[0,1])} \big|\,\Omega \models p - \Psi \models p\,\big| \;=\; \mathsf{kvd}\big(\Omega, \Psi\big).
\end{aligned}$$

(iii) We have to prove that for a (completely) positive unital map $f\colon A \to B$ between von Neumann algebras the associated state transformer $f_* = (-) \circ f\colon \mathrm{Hom}(B, \mathbb{C}) \to \mathrm{Hom}(A, \mathbb{C})$ is affine and non-expansive. The former is standard, so we concentrate on non-expansiveness. Let $\rho_1, \rho_2\colon B \to \mathbb{C}$ be states of B. Then:

$$\begin{aligned}
\mathsf{trd}\big(f_*(\rho_1), f_*(\rho_2)\big) &\stackrel{(8)}{=} \bigvee\nolimits_{e\in[0,1]_A} \big|\,f_*(\rho_1)(e) - f_*(\rho_2)(e)\,\big| \\
&= \bigvee\nolimits_{e\in[0,1]_A} \big|\,\rho_1(f(e)) - \rho_2(f(e))\,\big| \\
&\leq \bigvee\nolimits_{d\in[0,1]_B} \big|\,\rho_1(d) - \rho_2(d)\,\big| \stackrel{(8)}{=} \mathsf{trd}\big(\rho_1, \rho_2\big).
\end{aligned}$$ $\square$

Corollary 4.3 *The quantum state-and-effect triangle on the left below restricts to the triangle on the right.*

$$\begin{tikzcd}
\mathbf{EMod}^{\mathrm{op}} \arrow[rr, bend left] \arrow[rr, phantom, "\top"] & & \mathbf{Conv} \arrow[ll, bend left] \\
& \mathbf{vNA}^{\mathrm{op}} \arrow[ul, "\mathrm{Pred}"] \arrow[ur, "\mathrm{Stat}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbf{AEMod}^{\mathrm{op}} \arrow[rr, bend left] \arrow[rr, phantom, "\top"] & & \mathbf{ConvMet} \arrow[ll, bend left] \\
& \mathbf{vNA}^{\mathrm{op}} \arrow[ul, "\mathrm{Pred}"] \arrow[ur, "\mathrm{Stat}"'] &
\end{tikzcd} \tag{9}$$

Proof. It is standard that the self-adjoint elements of a von Neumann algebra A form an order unit space, and thereby that its predicates (effects) $[0,1]_A$ form an Archimedean effect module, see [13,14]. Hence the predicate functor $\mathrm{Pred} = [0,1]_{(-)}$ in (9) restricts to **AEMod**. The states functor Stat restricts by Lemma 4.2 (iii). The adjunction $\mathbf{AEMod}^{\mathrm{op}} \rightleftarrows \mathbf{ConvMet}$ was already established in (the proof of) Theorem 3.5. $\square$

5 Conclusions

We have used (total variation and trace) distances to get a better view on entwinedness of classical and quantum distributions, and on how they differ for some standard maximally entwined distributions. On a more general level, distances between classical and quantum states have been reformulated in logical terms and added to the state-and-effects triangles for classical and quantum probability.

References

[1] J. de Bakker and E. Vink. *Control Flow Semantics*. MIT Press, Cambridge, MA, 1996.

[2] P. Baldan, F. Bonchi, H. Kerstan, and B. König. Behavioral metrics via functor lifting. In V. Raman and S. Suresh, editors, *FSTTCS 2014: Foundations of Software Technology and Theoretical Computer Science*, volume 29 of *LIPIcs*, pages 403–415. Schloss Dagstuhl, 2014.

[3] F. van Breugel, C. Hermida, M. Makkai, and J. Worrell. Recursively defined metric spaces without contraction. *Theor. Comp. Sci.*, 380:143–163, 2007.

[4] K. Cho and B. Jacobs. The EfProb library for probabilistic calculations. In F. Bonchi and B. König, editors, *Conference on Algebra and Coalgebra in Computer Science (CALCO 2017)*, volume 72 of *LIPIcs*. Schloss Dagstuhl, 2017.

[5] K. Cho, B. Jacobs, A. Westerbaan, and B. Westerbaan. An introduction to effectus theory. see arxiv.org/abs/1512.05813, 2015.

[6] A. Gibbs and F. Su. On choosing and bounding probability metrics. *Int. Statistical Review*, 70(3):419–435, 2002.

[7] B. Jacobs. Probabilities, distribution monads, and convex categories. *Theor. Comp. Sci.*, 412(28):3323–3336, 2011.

[8] B. Jacobs. New directions in categorical logic, for classical, probabilistic and quantum logic. *Logical Methods in Comp. Sci.*, 11(3):1–76, 2015.

[9] B. Jacobs. A recipe for state and effect triangles. In L. Moss and P. Sobocinski, editors, *Conference on Algebra and Coalgebra in Computer Science (CALCO 2015)*, volume 35 of *LIPIcs*, pages 116–129. Schloss Dagstuhl, 2015.

[10] B. Jacobs. *Introduction to Coalgebra. Towards Mathematics of States and Observations*. Number 59 in Tracts in Theor. Comp. Sci. Cambridge Univ. Press, 2016.

[11] B. Jacobs. From probability monads to commutative effectuses. *Journ. of Logical and Algebraic Methods in Programming*, 156, 2017. See http://dx.doi.org/10.1016/j.jlamp.2016.11.006.

[12] B. Jacobs. A recipe for state and effect triangles. See https://arxiv.org/pdf/1703.09034, 2017.

[13] B. Jacobs and J. Mandemaker. The expectation monad in quantum foundations. In B. Jacobs, P. Selinger, and B. Spitters, editors, *Quantum Physics and Logic (QPL) 2011*, number 95 in Elect. Proc. in Theor. Comp. Sci., pages 143–182, 2012.

[14] B. Jacobs, J. Mandemaker, and R. Furber. The expectation monad in quantum foundations. Inf. & Comp., 2016.

[15] T. Lindvall. *Lectures on the Coupling Method*. John Wiley & Sons, New York, 1987.

[16] R. Madare, P. Panangaden, and G. Plotkin. Quantitative algebraic reasoning. In *Logic in Computer Science*. IEEE, Computer Science Press, 2016.

[17] A. McIver, C. Morgan, and T. Rabehaja. Abstract hidden Markov models: A monadic account of quantitative information flow. In *Logic in Computer Science*, pages 597–608. IEEE, Computer Science Press, 2015.

[18] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge Univ. Press, 2000.

[19] E. Rieffel and W. Polak. *Quantum Computing. A Gentle Introduction*. MIT Press, Cambridge, MA, 2011.

[20] C. Villani. *Optimal Transport — Old and New*. Springer, Berlin Heidelberg, 2009.

[21] N.S. Yanovsky and M.A. Manucci. *Quantum Computing for Computer Scientists*. Cambridge Univ. Press, 2008.