Entanglement in Sets and Vector Spaces

David Ellerman

UCR

April 2012

Comparison between set and quantum case

	Set Case	Quantum Case
Product	$X \times Y$	$H^A \otimes H^B$
Given state	$p(x,y)$	$\rho^{AB} = \|\psi\rangle\langle\psi\|$
Marginals	$p(x), p(y)$	ρ^A, ρ^B
Independent	$p(x,y) = p(x)\,p(y)$	$\rho^{AB} = \rho^A \otimes \rho^B$
Entangled	$p(x,y) \neq p(x)\,p(y)$	$\rho^{AB} \neq \rho^A \otimes \rho^B$
Bijection	$\{x_i\} \longleftrightarrow \{y_i\}$	$\{\|i_A\rangle\} \longleftrightarrow \{\|i_B\rangle\}$
Schmidt p_i	$p(x_i, y_i) = p_i$	$\|\psi\rangle = \sum_i \sqrt{p_i}\,\|i_A\rangle\,\|i_B\rangle$
Ent. Meas.	$d(p(x,y)\,\|\|\,p(x)\,p(y))$	$d\left(\rho^{AB}\|\|\rho^A \otimes \rho^B\right)$
Formula	$\sum_i p_i^2 - 2\sum_i p_i^3 + \sum_{i,j} p_i^2 p_j^2$	$1 - 2\sum_i p_i^3 + \sum_{i,j} p_i^2 p_j^2$
Max Entang.	$p_i = p_j$	$p_i = p_j$

- The research program of providing an objective indistinctness interpretation of QM uses, in part, the device of looking at the relevant mathematics of partitions at the level of sets, and then lifting the math to vector spaces where QM lives.
- Hence the beginning of understanding entanglement from this perspective is to understand it at the level of sets.
- Given a set X (thinking of the singletons as the set of completely distinct or eigen-elements), a subset S_X is the set-analogue of a superposition of its elements. Thus a block S_X in a partition on X is a "state" that is indistinct between its elements but has been distinguished from the elements outside of it.

What is entanglement?: II

- As more distinctions are made, the block gets refined eventually into singletons, the fully distinct or eigen-elements.
- To visualize this refinement of blocks, consider the powerset $\wp(X)$ partially ordered by inclusion and then flip it over (and throw away the null set) to get the *block refinement partial ordering* $\wp(X)^{op} - \emptyset$.
- Then you have the picture of the blob or ur-block X at the bottom and then all the other blocks that can result from making more and more distinctions until you arrive at the maximally distinguished atoms or singletons in the reverse-inclusion partial order.

What is entanglement?: III

- Each block or subset of X is understood as a mini-blob or objectively indistinct "element" which with more distinctions could be eventually refined into one its eigen-elements. It is indistinct between those eigen-alternatives, but it has been distinguished from all the other eigen-elements (outside the subset).
- Consider the reverse question of what happens if we go the other way of grouping the fully distinct eigen-elements together (i.e., superposing them) to get indistinct elements or blocks? Do we get anything new? In this case, no. All the blocks that can be obtained by forming new indistinct elements are just subsets of X already in the order $\wp(X)^{op} - \emptyset$.

What is entanglement?: IV

- Suppose we start with two sets X and Y and consider the two block-refinement orderings $\wp(X)^{op} - \emptyset$ and $\wp(Y)^{op} - \emptyset$. In each ordering by itself, nothing new appears if we superpose distinct eigen-elements to get blocks indistinct between their elements.
- Now we take the product of the two orderings $\left[\wp(X)^{op} - \emptyset\right] \times \left[\wp(Y)^{op} - \emptyset\right]$. The bottom of the ordering is $X \times Y$, the blob or ur-block in the combined system, and the fully distinct maximal elements in the product ordering are the eigen-pairs of singletons $(\{x\}, \{y\})$ or less pedantically $(x, y) \in X \times Y$, i.e., all the fully distinct alternatives that can be developed out of the ur-block $X \times Y$ by making distinctions.

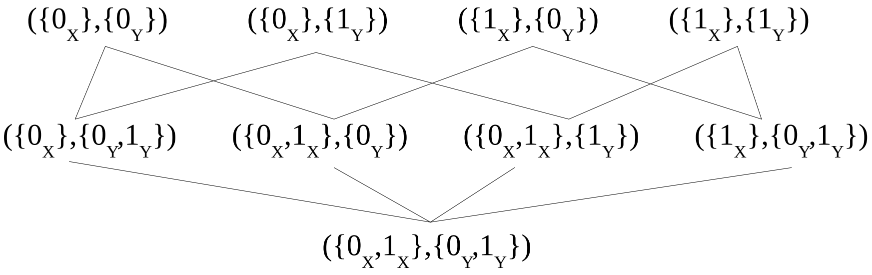
- Note that there is nothing new in the product eigen-elements; they are just pairs of eigen-elements from X and Y. The fully distinct eigen-elements of $\wp(X \times Y)^{op} - \emptyset$ are all there in $\left[\wp(X)^{op} - \emptyset\right] \times \left[\wp(Y)^{op} - \emptyset\right]$.
- We ask: if we group together or superpose some of these fully distinct eigen-pairs, $(\{x\}, \{y\})$ or simply (x, y) for $x \in X$ and $y \in Y$, do we get anything new or just the products of blocks of elements of X and of Y?
- We can get all the products of X-blocks and Y-blocks in this way, but we also get new indistinct blocks that are not product blocks.
- Those new blocks are the entangled states which represent the new ways to make indistinct elements due to the "interaction" of X and Y.

Example: I

- Consider the set version of two qubit space where $X = \{0_X, 1_X\}$ and $Y = \{0_Y, 1_Y\}$. The two block refinement orders are simple:

- Trivially, each superposition of the maximally distinct elements gives nothing new within each ordering.
- Then we take the product of the two orders to obtain:

Example: II



- A product block such as $(\{0_X\}, \{0_Y, 1_Y\})$ represents the subset $\{(0_X, 0_Y), (0_X, 1_Y)\} \subseteq X \times Y$ so this gives a suborder of the block refinement ordering for $X \times Y$.
- But not all subsets of $X \times Y$ can be obtained as products of blocks from X and Y.

- By "interacting" X and Y, new types of indistinct "states" can be formed by grouping together or "superposing" some of the fully distinct eigen-elements.
- For instance $\{(0_X, 0_Y), (1_X, 1_Y)\}$ and $\{(0_X, 1_Y), (1_X, 0_Y)\}$ are new types of "entangled" indistinct states as well as $\{(0_X, 0_Y), (0_X, 1_Y), (1_X, 1_Y)\}$ and three others (see continuation of the example below).
- This set example illustrates the **basic point** that by interacting two systems, no new fully distinct states are created but new types of indistinct states become possible and they are the entangled states. In spite of all the talk of entanglement as a uniquely quantum phenomenon, we see that it arises already at the level of ordinary sets. Now to measure it.

- Given two finite sets X and Y, a subset $S \subseteq X \times Y$ of their Cartesian product is a *product* subset if there are subsets $S_X \subseteq X$ and $S_Y \subseteq Y$ such that $S = S_X \times S_Y$.
- A subset $S \subseteq X \times Y$ that is not a product subset might be called an *entangled* subset.
- For any subset $S \subseteq X \times Y$, a natural measure of its entanglement can be constructed by first viewing S as the support of the equiprobable or Laplacian joint probability distribution on S.
- If $|S| = N$, then define $p(x,y) = \frac{1}{N}$ if $(x,y) \in S$ and $p(x,y) = 0$ otherwise.
- The marginal distributions are defined in the usual way:
  - $p(x) = \sum_y p(x,y)$

Entanglement of sets: II

- $p(y) = \sum_x p(x, y)$.

- A joint probability distribution $p(x, y)$ on $X \times Y$ is *independent* if for all $(x, y) \in X \times Y$,

$$p(x, y) = p(x)\, p(y).$$

Independent distribution

Theorem

A subset $S \subseteq X \times Y$ is entangled iff the equiprobable distribution on S is not independent.

Proof: Let S_X be the support or projection of S on X, i.e., $S_X = \{x : \exists y \in Y, (x, y) \in S\}$ and similarly for S_Y. If S is not entangled, i.e., $S = S_X \times S_Y$, then $p(x) = |S_Y| / N$ for $x \in S_X$ and

$p(y) = |S_X|/N$ for $y \in S_Y$ where $|S_X||S_Y| = N$. Then for $(x, y) \in S$,

$$p(x, y) = \frac{1}{N} = \frac{N}{N^2} = \frac{|S_X||S_Y|}{N^2} = p(x)\,p(y)$$

and $p(x, y) = 0 = p(x)\,p(y)$ for $(x, y) \notin S$ so the equiprobable distribution is independent. If $S \neq S_X \times S_Y$, then $S \subsetneqq S_X \times S_Y$ so let $(x, y) \in S_X \times S_Y - S$. Then $p(x), p(y) > 0$ but $p(x, y) = 0$ so it is not independent. $\square$

- Hence for sets, a measure of entanglement of a subset $S \subseteq X \times Y$ would the measure of the logical divergence between the equiprobable distribution $p(x, y)$ on S and the product of marginals distribution $p(x)\,p(y)$:

$$d\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) =$$
$$2h\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) - h\left(p\left(x,y\right)\right) - h\left(p\left(x\right)p\left(y\right)\right).$$

- $d\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) > 0$ iff S is an entangled subset and
- $d\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) = 0$ iff S is a product subset.
- The logical entropy of the equiprobable distribution is $h\left(p\left(x,y\right)\right) = 1 - N\frac{1}{N^2} = 1 - \frac{1}{N}$ where $|S| = N$.
- For any $x \in S_X$, let $s_x = |\{y : (x,y) \in S\}|$ and similarly for s_y. Then $p\left(x\right) = \frac{s_x}{N}$ and $p\left(y\right) = \frac{s_y}{N}$ so the logical entropy of the product distribution is

$$h\left(p\left(x\right)p\left(y\right)\right) = 1 - \sum_{x \in S_X, y \in S_Y} \frac{s_x^2 s_y^2}{N^4}.$$

Entanglement of sets: V

- The cross-entropy is:

$$h\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) = 1 - \sum_{(x,y)\in S} \frac{s_x s_y}{N^3}.$$

- Hence putting it together:

$$d\left(p||p\left(x\right)p\left(y\right)\right) = 2h\left(p||p\left(x\right)p\left(y\right)\right) - h\left(p\right) - h\left(p\left(x\right)p\left(y\right)\right)$$

$$= 2\left[1 - \sum_{(x,y)\in S} \frac{s_x s_y}{N^3}\right] - \left[1 - \frac{1}{N}\right] - \left[1 - \sum_{x\in S_X, y\in S_Y} \frac{s_x^2 s_y^2}{N^4}\right]$$

$$\boxed{d\left(p||p\left(x\right)p\left(y\right)\right) = \sum_{x\in S_X, y\in S_Y} \frac{s_x^2 s_y^2}{N^4} + \frac{1}{N} - 2\sum_{(x,y)\in S} \frac{s_x s_y}{N^3}}.$$

Entanglement of sets: VI

- In the relevant range, $1 \leq s_x \leq |S_X|$ and $1 \leq s_y \leq |S_Y|$, the divergence is inversely related to the s_x and s_y so the maximum entanglement (by this measure) occurs when $s_x = s_y = 1$ which means that S is the graph of a bijection between a subset of X and a subset of Y (which in combinatorics is a partial matching between X and Y).
- In that maximum entanglement case, the value of the divergence is:

$$MaxDiv = \sum_{x \in S_X, y \in S_Y} \frac{s_x^2 s_y^2}{N^4} + \frac{1}{N} - 2\sum_{(x,y) \in S} \frac{s_x s_y}{N^3}$$
$$= |S_X|\,|S_Y|\,\frac{1}{N^4} + \frac{1}{N} - 2\frac{1}{N^2} = \frac{1}{N^4}\left[|S_X|\,|S_Y| + N^3 - 2N^2\right].$$

Since S is the graph of a bijection between a subset of X and of Y, we might as well throw away the rest of X and Y so that $|S_X| = |S_Y| = N$, and then we have:

Entanglement of sets: VII

$$MaxDiv = \frac{1}{N^4}\left[|S_X|\,|S_Y| + N^3 - 2N^2\right] = \frac{1}{N^4}\left[N^3 - N^2\right] \text{ so}$$
$$\boxed{MaxDiv = \frac{1}{N}\left[1 - \frac{1}{N}\right]}.$$

- Alternatively, instead of cutting down X and Y so that $|S_X| = |S_Y| = N$, we can increase the *MaxDiv* by increasing N (since the derivative of $\frac{1}{N}\left[1 - \frac{1}{N}\right]$ is positive for positive N) until $N = \min\left(|X|, |Y|\right)$. Then we could throw away the excess in X or Y and use the above *MaxDiv* formula where a bijection, that is a *complete* matching between X and Y, is the maximum entanglement subset.

Example continued: I

- Consider again the set version of two qubit space where $X = \{0_X, 1_X\}$ and $Y = \{0_Y, 1_Y\}$.
- The product space $X \times Y$ has 15 nonempty subsets. Each factor X and Y has 3 nonempty subsets so $3 \times 3 = 9$ of the 15 subsets are product subsets leaving 6 entangled subsets.
- The 6 entangled subsets and their divergences are:

$\{(0,0),(1,1)\}$	$\{(0,1),(1,0)\}$
$\frac{1}{4}$	$\frac{1}{4}$
$\{(0,0),(0,1),(1,0)\}$	$\{(0,0),(0,1),(1,1)\}$
$\frac{4}{81}$	$\frac{4}{81}$
$\{(0,1),(1,0),(1,1)\}$	$\{(0,0),(1,0),(1,1)\}$
$\frac{4}{81}$	$\frac{4}{81}$

Example continued: II

- The first two "Bell subsets" are the two graphs of bijections $X \longleftrightarrow Y$ and have the maximum entanglement which can be calculated by the formula for the maximum divergence where $N = 2$, $MaxDiv = \frac{1}{2}\left[1 - \frac{1}{2}\right] = \frac{1}{4}$.
- The entanglement of say $\{(0,0), (0,1), (1,0)\}$ is calculated using $N = 3$, $s_{0X} = 2 = s_{0Y}$ and $s_{1X} = 1 = s_{1Y}$.
- All the 9 product states have zero entanglement. For instance, for $S = \{(0,0), (0,1)\}$, we have $N = 2$, $s_{0X} = 2$, $s_{1X} = 0$, and $s_{0Y} = s_{1Y} = 1$ so that:

$$d\left(p||p\left(x\right)p\left(y\right)\right) = \sum_{x \in S_X, y \in S_Y} \frac{s_x^2 s_y^2}{N^4} + \frac{1}{N} - 2\sum_{(x,y) \in S} \frac{s_x s_y}{N^3}$$
$$= \left[\frac{4}{16} + \frac{4}{16}\right] + \frac{1}{2} - 2\left[\frac{2}{8} + \frac{2}{8}\right] = \frac{1}{2} + \frac{1}{2} - 1 = 0.$$

Digression on probabilities as random variables: I

- Any finite probability distribution $p = \{p_1, ..., p_n\}$ can be viewed as a random variable taking the value p_i with the probability p_i.
- The *expectation* of p is $E_p(p) = \sum_i p_i^2$ so the logical entropy $h(p) = 1 - \sum_i p_i^2$ is the complement of the expectation of p.
- Given another distribution $q = \{q_1, ..., q_n\}$ over the same index set, the *cross-expectation* is:

$$E(p||q) = E_p(q) = E_q(p) = \sum_i p_i q_i$$

- The logical cross-entropy $h(p||q) = 1 - \sum_i p_i q_i$ is the complement of the cross-expectation.

Digression on probabilities as random variables: II

- The logical divergence is:

$$d(p||q) = \sum_i (p_i - q_i)^2 = 2h(p||q) - h(p) - h(q)$$
$$= 2[1 - E(p||q)] - [1 - E_p(p)] - [1 - E_q(q)]$$
$$= E_p(p) + E_q(q) - 2E(p||q) = E_p(p) + E_q(q) - E_p(q) - E_q(p)$$
$$\boxed{d(p||q) = (E_p - E_q)(p - q)}$$

Divergence in terms of linear expectation operators

- The logical information inequality that $d(p||q) \geq 0$ can then be written as:

$$E_p(p) + E_q(q) \geq E_p(q) + E_q(p)$$

Sum of self-expectations $\geq$ sum of cross-expectations.

Probabilities on bijections: I

- In the set case, a subset $S \subseteq X \times Y$ that is the graph of a bijection is the set analogue of the Schmidt decomposition of a pure state on a tensor product which is always available when working over Hilbert spaces. The different pairs of orthogonal basis states in a Schmidt decomposition $|\psi\rangle = \sum_i \sqrt{p_i}\,|i_A\rangle \otimes |i_B\rangle$ may have different Schmidt coefficients $\sqrt{p_i}$. Hence to develop the set analogue, we assume S is a bijection graph but allow an arbitrary probability distribution on S.
- A bijective S has the form $\{(x_i, y_i) : i = 0, ..., N-1\}$ so we assume a probability distribution with $p(x_i, y_i) = p_i$ and 0 otherwise.

Probabilities on bijections: II

- Since the set S is a bijection, the marginals are $p(x_i) = p_i = p(y_i)$, so that $h(p(x)) = h(p(y)) = 1 - \sum_i p_i^2$. This is the set analogue of the reduced density matrices having the same eigenvalues p_i in the quantum case.
- The logical entropy of $p(x,y)$ is: $h(p(x,y)) = 1 - \sum_i p_i^2$.
- The logical entropy of the product distribution is: $h(p(x)p(y)) = 1 - \sum_{i,j} p_i^2 p_j^2 = 1 - \left(\sum_i p_i^2\right)^2$.
- The cross-entropy is: $h(p(x,y)||p(x)p(y)) = 1 - \sum_i p_i^3$.
- Hence the measure of entanglement in this case is:

Probabilities on bijections: III

$$d\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) =$$
$$2h\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) - h\left(p\left(x,y\right)\right) - h\left(p\left(x\right)p\left(y\right)\right)$$
$$= 2\left[1 - \sum_i p_i^3\right] - \left[1 - \sum_i p_i^2\right] - \left[1 - \left(\sum_i p_i^2\right)^2\right] \text{ so}$$
$$\boxed{d\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) = \sum_i p_i^2 - 2\sum_i p_i^3 + \left(\sum_i p_i^2\right)^2}.$$

- In terms of the expectations,
  - $E_p\left(p\right) = \sum_i p_i^2$,
  - $E_{p(x)p(y)}\left(p\left(x\right)p\left(y\right)\right) = \sum_{i,j} p_i^2 p_j^2 = \left(\sum_i p_i^2\right)^2$,
  - $E\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) = \sum_i p_i^3$.
- The variance of the random variable p is:

$$Var\left(p\right) = E_p\left(p^2\right) - E_p\left(p\right)^2 = \sum_i p_i^3 - \left(\sum_i p_i^2\right)^2.$$

- Hence the divergence formula in this special case is:

$$d\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) = E_p\left[p\left(x,y\right) - p\left(x\right)p\left(y\right)\right] - Var\left(p\right).$$

- Here again, the maximum divergence & entanglement is the equiprobable case, $p_i = \frac{1}{N}$, where $E_p\left(p\right) = \frac{1}{N}$, $E_p\left(p\left(x\right)p\left(y\right)\right) = \frac{1}{N^2}$, and $Var\left(p\right) = 0$ so the formula gives the previous result:

$$\frac{1}{N} - \frac{1}{N^2} = \frac{1}{N}\left[1 - \frac{1}{N}\right].$$

- Note how the variance of p takes away from the divergence in this bijective case, and the variance is 0 for both the extreme cases: $p_i = \frac{1}{N}$ and $p_1 = 1$. The first case ($p_i = \frac{1}{N}$) is the maximum entanglement and the second case ($p_1 = 1$) is zero entanglement (product state).

Quantum case: I

- Given two systems A, B represented in Hilbert spaces H^A and H^B, let $\rho^{AB} = |\psi\rangle\langle\psi|$ be a *pure* state in the tensor product $H^A \otimes H^B$.
- If $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$ are orthonormal bases for the component spaces, let $[\alpha]$ be the matrix of coefficients for ψ, i.e.,

$$[\alpha] = [\alpha_{ij}] \text{ where } |\psi\rangle = \sum_{i,j} \alpha_{ij} |a_i\rangle \otimes |b_j\rangle.$$

- Then $[\alpha][\alpha]^\dagger = \rho^A$ is the reduced density matrix on H^A and $[\alpha]^\dagger [\alpha] = \rho^B$ is the reduced density matrix on H^B, where $[]^\dagger$ is the Hermitian transpose.

Quantum case: II

- The Schmidt decomposition of ψ is $|\psi\rangle = \sum_i \sqrt{p_i}\,|i_A\rangle \otimes |i_B\rangle$ where $\{|i_A\rangle\}$ and $\{|i_B\rangle\}$ are orthonormal bases for the component spaces.
- Where $p_i \neq 0$, the Schmidt decomposition establishes a bijection between a subset of the basis $\{|i_A\rangle\}$ and a subset of the basis $\{|i_B\rangle\}$ which is the "lift" of such a bijection in the set case.
- The reduced density matrices can then be expressed as: $\rho^A = \sum_i p_i\,|i_A\rangle\,\langle i_A|$ and $\rho^B = \sum_i p_i\,|i_B\rangle\,\langle i_B|$ so the p_i are the non-negative eigenvalues for both reduced density matrices.
- Since the trace is invariant under similarity transformations and since each density matrix could be diagonalized to its diagonal matrix of eigenvalues, the traces of the squares are:

$$\mathrm{tr}\left[\left(\rho^A\right)^2\right] = \mathrm{tr}\left[\left(\rho^B\right)^2\right] = \sum_i p_i^2$$

so that $h\left(\rho^A\right) = h\left(\rho^B\right) = 1 - \sum_i p_i^2$.

- Since ρ^{AB} is assumed to be a pure state, $\mathrm{tr}\left[\left(\rho^{AB}\right)^2\right] = 1$ so $h\left(\rho^{AB}\right) = 1 - \mathrm{tr}\left[\left(\rho^{AB}\right)^2\right] = 0$.
- The logical entropy of the product state $\rho^A \otimes \rho^B$ is:

$$h\left(\rho^A \otimes \rho^B\right) = 1 - \mathrm{tr}\left[\left(\rho^A\right)^2\right]\mathrm{tr}\left[\left(\rho^B\right)^2\right] = 1 - \left(\sum_i p_i^2\right)^2.$$

Quantum case: IV

- The Schmidt number is the number of non-zero p_i, and it is 1 with $p_1 = 1$ iff ρ^{AB} is a product state, i.e., $\rho^{AB} = \rho^A \otimes \rho^B$. Then $\sum_i p_i^2 = 1$ and $h\left(\rho^A\right) = h\left(\rho^B\right) = 0$ so that $d\left(\rho^{AB} || \rho^A \otimes \rho^B\right) = 0$ as well.
- In the Schmidt basis for the case where both H^A and H^B are three dimensional, then:

$$|\psi\rangle = \sqrt{p_0}\,|0_A\rangle \otimes |0_B\rangle + \sqrt{p_1}\,|1_A\rangle \otimes |1_B\rangle + \sqrt{p_2}\,|2_A\rangle \otimes |2_B\rangle.$$

- Then the matrix for ρ^{AB} in the $\{|i_A\rangle \otimes |j_B\rangle\}$ basis is:

Quantum case: V

$$\rho^{AB} = \begin{bmatrix} p_0 & 0 & 0 & 0 & \sqrt{p_0 p_1} & 0 & 0 & 0 & \sqrt{p_0 p_2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sqrt{p_1 p_0} & 0 & 0 & 0 & p_1 & 0 & 0 & 0 & \sqrt{p_1 p_2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sqrt{p_2 p_0} & 0 & 0 & 0 & \sqrt{p_2 p_1} & 0 & 0 & 0 & p_2 \end{bmatrix}.$$

- The matrix for $\rho^A \otimes \rho^B$ is diagonal:

$$\rho^A \otimes \rho^B = \begin{bmatrix} p_0^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_0p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_0p_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_1p_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & p_1p_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_2p_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_2p_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_2^2 \end{bmatrix}.$$

- The cross-entropy $h\left(\rho^{AB}||\rho^A \otimes \rho^B\right) = 1 - \operatorname{tr}\left[\rho^{AB}\left(\rho^A \otimes \rho^B\right)\right]$ will in general just pick out the cubic terms:

$$h\left(\rho^{AB}||\rho^A \otimes \rho^B\right) = 1 - \sum_i p_i^3.$$

- The suggested measure of entanglement is the logical divergence $d\left(\rho^{AB}||\rho^{A}\otimes\rho^{B}\right)$ which can be computed as:

$$\begin{aligned} d\left(\rho^{AB}||\rho^{A}\otimes\rho^{B}\right) &= 2h\left(\rho^{AB}||\rho^{A}\otimes\rho^{B}\right) - h\left(\rho^{AB}\right) - h\left(\rho^{A}\otimes\rho^{B}\right) \\ &= 2\left[1-\operatorname{tr}\left[\rho^{AB}\left(\rho^{A}\otimes\rho^{B}\right)\right]\right] - \left[1-\operatorname{tr}\left[\left(\rho^{AB}\right)^{2}\right]\right] - \\ &\left[1-\operatorname{tr}\left[\left(\rho^{A}\otimes\rho^{B}\right)^{2}\right]\right] \\ &= \operatorname{tr}\left[\left(\rho^{AB}\right)^{2}\right] - 2\operatorname{tr}\left[\rho^{AB}\left(\rho^{A}\otimes\rho^{B}\right)\right] + \operatorname{tr}\left[\left(\rho^{A}\otimes\rho^{B}\right)^{2}\right] \end{aligned}$$

which is the lift of the set version:

$$d\left(p\left(x,y\right)||p\left(x\right)p\left(y\right)\right) = \sum_i p_i^2 - 2\sum_i p_i^3 + \sum_{i,j} p_i^2 p_j^2.$$

- Since we have assumed that ρ^{AB} is a pure state (in order to use the Schmidt decomposition), $\mathrm{tr}\left[\left(\rho^{AB}\right)^2\right] = 1$ so the final formula for the entanglement measure in terms of the Schmidt coefficients is:

$$\boxed{d\left(\rho^{AB}||\rho^{A}\otimes\rho^{B}\right) = 1 - 2\sum_i p_i^3 + \left(\sum_i p_i^2\right)^2}.$$

Entanglement measure for pure ρ^{AB}
in terms of Schmidt coefficients

- Here again, it can be shown that the entanglement measure is a maximum when the non-zero Schmidt coefficients are equal so if there are n non-zero p_i's, then $p_1 = ... = p_n = \frac{1}{n}$. Such a case is often called "maximally entangled" so the entanglement measure agrees.

- When all the Schmidt coefficients are equal, the value of the maximum divergence is:

$$d\left(\rho^{AB}||\rho^{A}\otimes\rho^{B}\right)=1-2\sum_i p_i^3+\left(\sum_i p_i^2\right)^2$$
$$=1-2\frac{n}{n^3}+\left(\frac{n}{n^2}\right)^2=1-2\frac{n^2}{n^4}+\frac{n^2}{n^4}=1-\frac{n^2}{n^4}=1-\frac{1}{n^2}$$

which differs from the set formula in the first term which is $\mathrm{tr}\left[\left(\rho^{AB}\right)^2\right]=1$ instead of $\sum_i p_i^2$ since there is no set analogue of a non-trivial pure state. In the set case, a "pure state" is the trivial case $p_1=1$ and then indeed $\sum_i p_i^2=1$.

- The Schmidt coefficients (squared) give a probability distribution $p=\{p_1,...,p_n\}$ so we may restate the divergence formula using the expectations and variance:

$$d\left(\rho^{AB}||\rho^{A}\otimes\rho^{B}\right)=1-\sum_i p_i^3-Var\left(p\right)=$$
$$h\left(\rho^{AB}||\rho^{A}\otimes\rho^{B}\right)-Var\left(p\right)$$

where the variance in the Schmidt coefficients is 0 in both the extreme cases of maximum entanglement and zero entanglement.

- For all the Bell basis vectors in two qubit space, $p_1 = p_2 = \frac{1}{2}$ and their maximal measure of entanglement is $1-\left(\frac{1}{2}\right)^2=\frac{3}{4}$.

Comparison between set and quantum case

	Set Case	Quantum Case
Product	$X \times Y$	$H^A \otimes H^B$
Given state	$p(x,y)$	$\rho^{AB} = \vert\psi\rangle\langle\psi\vert$
Marginals	$p(x), p(y)$	ρ^A, ρ^B
Independent	$p(x,y) = p(x)\,p(y)$	$\rho^{AB} = \rho^A \otimes \rho^B$
Entangled	$p(x,y) \neq p(x)\,p(y)$	$\rho^{AB} \neq \rho^A \otimes \rho^B$
Bijection	$\{x_i\} \longleftrightarrow \{y_i\}$	$\{\vert i_A\rangle\} \longleftrightarrow \{\vert i_B\rangle\}$
Schmidt p_i	$p(x_i, y_i) = p_i$	$\vert\psi\rangle = \sum_i \sqrt{p_i}\,\vert i_A\rangle\,\vert i_B\rangle$
Ent. Meas.	$d(p(x,y)\,\Vert\,p(x)\,p(y))$	$d\left(\rho^{AB}\Vert\rho^A \otimes \rho^B\right)$
Formula	$\sum_i p_i^2 - 2\sum_i p_i^3 + \sum_{i,j} p_i^2 p_j^2$	$1 - 2\sum_i p_i^3 + \sum_{i,j} p_i^2 p_j^2$
Max Entang.	$p_i = p_j$	$p_i = p_j$

De-lifting Density Matrices

David Ellerman

UCR

May 2012

De-lifting table to be explained

Quantum case	Set case
$\rho = \sum_k q_k \lvert\psi_k\rangle \langle\psi_k\rvert$ (orth.de.)	$p_\pi = \sum_k p_{B_k} \lvert\sqrt{p_{i\lvert B_k}}\rangle\langle\sqrt{p_{i\lvert B_k}}\rvert$
Partition of eigenspaces $\{E_k\}$	Partition of blocks $\pi = \{B_k\}$
Orthonormal basis $\{\lvert i\rangle\}$	Set of points $U = \{i\}$
$\{p_i\}$ pdf associated with $\{\lvert i\rangle\}$	Pdf $\{p_i\}$ for $i \in U$
$\{q_k\}$ pdf associated with $\{E_k\}$	$\{p_B\}$ pdf associated with $\{B_k\}$
Pure states $\lvert\psi_k\rangle \langle\psi_k\rvert$	Pure states $\lvert\sqrt{p_{i\lvert B_k}}\rangle\langle\sqrt{p_{i\lvert B_k}}\rvert$
1-dimensional eigenspaces E_k	Discrete partition $\pi = 1$
Log. entropy $h(\rho) = 1 - \mathrm{tr}\left[\rho^2\right]$	Log. entropy $h(p_\pi) = 1 - \mathrm{tr}\left[p_\pi^2\right]$
$\left\lvert\rho_{ij}\right\rvert^2 =$ indit probability	$(p_\pi)_{ij}^2 =$ indit probability
$h(\rho) = \Sigma$ dit probabilities	$h(p_\pi) = \Sigma$ dit probabilities
$h(\hat{\rho}) - h(\rho) = \Sigma$ new dit probs.	$h(p_{\hat{\pi}}) - h(p_\pi) = \Sigma$ new dit probs.
$\overline{T} = \mathrm{tr}\left[T\rho\right]$	$\overline{T} = \mathrm{tr}\left[Tp_\pi\right]$

Lifting and de-lifting as an engine of understanding: I

- We have seen how lifting a set concept to a vector space concept can be a way to "understand" the vector space concept.
- For instance, there is the "mystery" as to why numerical attributes like momentum in classical physics become linear operators in QM.
- For sets, a numerical attribute is a function $f : U \to \mathbb{R}$ from the set into some field such as $\mathbb{R}$.
- We can express f as the sum of its values r times the characteristic functions for the subsets $f^{-1}(r) \subseteq U$ where f takes that value r:

Lifting and de-lifting as an engine of understanding: II

$$f = \sum_r r\chi_{f^{-1}(r)}.$$

"Spectral decomposition" of numerical attribute

- Then we do the lift:
  - Values r lift to eigenvalues λ;
  - Subset where f takes a value r is $f^{-1}(r)$ lifts to eigenspace W_λ for the eigenvalue λ;
  - Characteristic function $\chi_{f^{-1}(r)}$ for $f^{-1}(r)$ lifts to projection operator P_λ to eigenspace W_λ.
  - Check that characteristic functions idempotently *multiply*, i.e., $\chi_S^2(u) = \chi_S(u)\chi_S(u) = \chi_S(u)$, just as projection operators idempotently *compose*, i.e., $P_W^2 = P_W \cdot P_W = P_W$.✓

Lifting and de-lifting as an engine of understanding: III

- "Spectral decomposition" of (total) numerical attribute $f = \sum_r r\chi_{f^{-1}(r)}$ lifts to spectral decomposition $L = \sum_\lambda \lambda P_\lambda$ of (diagonalizable) linear operator L.

- In this manner, we *understand* why a quantum observable, that corresponds to a classical numerical attribute, is a linear operator where the eigenvalues are the values of the quantum attribute (and since the values need to be real, the operator is Hermitian).
- Thus, by the lifting program, we essentially *derive* that a real-valued quantum attribute is given by a Hermitian linear operator on the state space.

- A general formulation of QM works not just with state vectors but with the notion of density operators, which become density matrices when represented in a basis.
- Density matrices can be better understood by de-lifting that vector-space concept to obtain the corresponding set-concept.
- For any density operator ρ, we first represent it using its *orthogonal decomposition*: $\rho = \sum_k q_k |\psi_k\rangle \langle\psi_k|$ where the pure state density operators (i.e., projection operators) $|\psi_k\rangle \langle\psi_k|$ have orthogonal support since they project to the orthogonal eigenspaces E_k of ρ (where q_k are the non-negative eigenvalues that sum to 1).

De-lifting density matrices: II

- We can choose an orthonormal basis for each eigenspace E_k. The union of these disjoint eigenspace bases form an orthonormal basis $\{|i\rangle\}$ (for $i = 0, 1, ..., n-1$) for the whole space (since the eigenspaces span the space).
- When the projection operator $|\psi_k\rangle\langle\psi_k|$ to the eigenspace E_k is represented in the chosen basis for that eigenspace, then the diagonal entries in the density matrix for $|\psi_k\rangle\langle\psi_k|$ are a probability distribution with a probability associated with each basis vector for E_k.
- When those probabilities associated with the bases for the E_k are weighted by the probabilities q_k, then we have a probability distribution $\{p_i\}$ associated with the basis $\{|i\rangle\}$ for the whole space.

De-lifting density matrices: III

- Now we de-lift to obtain the set version of a density operator $\rho = \sum_k q_k |\psi_k\rangle \langle\psi_k|$:
  - The orthonormal basis $\{|i\rangle\}_{i=0,\ldots,n-1}$ for the space de-lifts to a set $U = \{0, 1, ..., n-1\}$ of clear and distinct points.
  - The vector space partition of the eigenspaces E_k is generated by the set-partition of $\{|i\rangle\}$ whose blocks generate the eigenspaces E_k, so it de-lifts to a set-partition $\pi = \{B_k\}$ of U.
  - The probabilities $\{p_i\}$ associated with the basis set $\{|i\rangle\}$ de-lifts to a probability distribution $\{p_i\}$ over the points of $U = \{0, 1, ..., n-1\}$.
- For expository purposes, the simplest case to indicate the general pattern is $U = \{0, 1, 2\}$.

De-lifting density matrices: IV

- If $\pi = \mathbf{0} = \{U\}$, the indiscrete partition on U, then the *density matrix representation* of $p_{\mathbf{0}}$ is the $n \times n$ matrix with the ij-entry $\sqrt{p_i p_j}$. If we take $\left|\sqrt{p}\right\rangle$ as the column vector of the square roots $\sqrt{p_i}$ and $\left\langle\sqrt{p}\right|$ as its transpose, then the density matrix is $p_{\mathbf{0}} = \left|\sqrt{p}\right\rangle\left\langle\sqrt{p}\right|$. In the case of $n = 3$, this is:

$$p_{\mathbf{0}} = \left|\sqrt{p}\right\rangle\left\langle\sqrt{p}\right| = \begin{bmatrix} \sqrt{p_0} \\ \sqrt{p_1} \\ \sqrt{p_2} \end{bmatrix} \begin{bmatrix} \sqrt{p_0} & \sqrt{p_1} & \sqrt{p_2} \end{bmatrix}$$
$$= \begin{bmatrix} p_0 & \sqrt{p_0 p_1} & \sqrt{p_0 p_2} \\ \sqrt{p_0 p_1} & p_1 & \sqrt{p_1 p_2} \\ \sqrt{p_0 p_2} & \sqrt{p_1 p_2} & p_2 \end{bmatrix}.$$

De-lifting density matrices: V

- The density matrix representation of $p_{\mathbf{0}}$ is a symmetric, positive semi-definite matrix of trace 1. Since there is only one block in the partition $\mathbf{0}$, $p_{\mathbf{0}}$ is a pure state density matrix, i.e., it is a projection matrix.
- As in the quantum case for a pure state $\rho = 1\left|\psi\right\rangle\left\langle\psi\right|$, the trace of this "pure state" matrix squared is 1:

$$\mathbf{tr}\left[p_{\mathbf{0}}^2\right] = p_0^2 + p_0p_1 + p_0p_2 + p_1^2 + p_0p_1 + p_1p_2 + p_2^2 + p_0p_2 + p_1p_2$$
$$= p_0\left(p_0 + p_1 + p_2\right) + p_1\left(p_0 + p_1 + p_2\right) + p_2\left(p_0 + p_1 + p_2\right) = 1.$$

Interpreting density matrix entries: I

- Let us define an *amplitude for a probabilistic event* as a quantity whose absolute-value squared or "absolute square" is the probability of the event.
- An outcome i in the sample space U has probability p_i. If we take two independent drawings from the sample space (w/replacement), then the probability of the pair (i, j) of outcomes is $p_i p_j$.
- Thus $\sqrt{p_i p_j}$ is an amplitude for the pair-outcome (i, j) and p_i is an amplitude for the pair-outcome (i, i). Hence we have:

$$\rho_{\mathbf{0}} = \begin{bmatrix} p_0 & \sqrt{p_0 p_1} & \sqrt{p_0 p_2} \\ \sqrt{p_0 p_1} & p_1 & \sqrt{p_1 p_2} \\ \sqrt{p_0 p_2} & \sqrt{p_1 p_2} & p_2 \end{bmatrix}$$

= matrix of pair-outcome amplitudes.

Interpreting density matrix entries: II

- For comparison purposes, let's recall the quantum case of a pure state density matrix.
- Consider any *pure* state $\rho = |\psi\rangle\langle\psi|$. The general three-dimensional case is illustrative:

$$|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle + \alpha_2 |2\rangle$$
$$\rho = |\psi\rangle\langle\psi| = [\alpha_0 |0\rangle + \alpha_1 |1\rangle + \alpha_2 |2\rangle]\left[\alpha_0^* \langle 0| + \alpha_1^* \langle 1| + \alpha_2^* \langle 2|\right]$$
$$= \begin{bmatrix} \alpha_0\alpha_0^* & \alpha_0\alpha_1^* & \alpha_0\alpha_2^* \\ \alpha_1\alpha_0^* & \alpha_1\alpha_1^* & \alpha_1\alpha_2^* \\ \alpha_2\alpha_0^* & \alpha_2\alpha_1^* & \alpha_2\alpha_2^* \end{bmatrix}.$$

Interpreting density matrix entries: III

- A diagonal term $\rho_i = \alpha_i \alpha_i^*$ is the probability that a $\{|i\rangle\}$-basis measurement of the state $|\psi\rangle$ will result in the eigenstate $|i\rangle$. Hence ρ_i squared is the probability for the pair-outcome of getting $(|i\rangle, |i\rangle)$ in two independent measurements.
- The probability $\rho_i \rho_j = \alpha_i \alpha_i^* \alpha_j \alpha_j^*$ is the probability that two independent measurements would result in the pair of eigenstates $(|i\rangle, |j\rangle)$.
- The off-diagonal *coherence term* ρ_{ij} of ρ is the amplitude $\alpha_i \alpha_j^* = \langle i|\psi\rangle \langle \psi|j\rangle$ whose corresponding probability is:

$$\left|\rho_{ij}\right|^2 = \rho_{ij}\rho_{ji} = \alpha_i \alpha_j^* \alpha_j \alpha_i^* = \alpha_i \alpha_i^* \alpha_j \alpha_j^* = \rho_i \rho_j$$

Probability of two measurements giving pair $(|i\rangle, |j\rangle)$.

Interpreting density matrix entries: IV

- Hence the pure state density matrix ρ is the matrix of amplitudes ρ_{ij} whose absolute squares are the probabilities for the pair-outcomes $(|i\rangle, |j\rangle)$ in two independent measurements.

Density matrix for a set-partition with probabilities: I

- The meaning of a "pure state" is clarified by the more general density matrix representation of a set-partition $\pi = \{B\}$ on U with point probabilities p_i for $i \in U$.
- In this case of a set-partition $\pi = \{B\}$ on U, with probabilities assigned to the elements of U (e.g., the Laplacian assumption of equal probabilities), then we can sum the probabilities of the elements in a block B to arrive at a block-probability $p_B = \sum_{i \in B} p_i$. The set of block probabilities $\{p_B\}_{B \in \pi}$ is also a probability distribution (the de-lift of the $\{q_k\}$ distribution of probabilities associated with the eigenspace "blocks" E_k).

Density matrix for a set-partition with probabilities: II

- For each block B, there is the conditional probability distribution that can be viewed as a "pure state" density matrix where the probabilities of the points are:

$$p_{i|B} = \Pr(i|B) = \begin{cases} \frac{p_i}{p_B} \text{ if } i \in B \\ 0 \text{ if } i \notin B \end{cases}.$$

- With a reordering of points, this "pure state" density matrix $\left|\sqrt{p_{i|B}}\right\rangle\left\langle\sqrt{p_{i|B}}\right|$ has a matrix-block corresponding to the partition block (which is the de-lift of the density matrix for the pure state projection operator $|\psi_k\rangle\langle\psi_k|$ for the "block" E_k in the vector space partition of eigenspaces);

Density matrix for a set-partition with probabilities: III

$$\left|\sqrt{p_{i|B}}\right\rangle\left\langle\sqrt{p_{i|B}}\right| = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{p_i}{p_B} & \cdots & \frac{\sqrt{p_i p_j}}{p_B} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \frac{\sqrt{p_i p_j}}{p_B} & \cdots & \frac{p_j}{p_B} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

- Then the natural density matrix representation p_π of the partition π with the given point probabilities p_i and the block probabilities $\{p_B\}$ is the *mixed state density matrix*:

Density matrix for a set-partition with probabilities: IV

$$p_\pi = \sum_{B\in\pi} p_B \left|\sqrt{p_{i|B}}\right\rangle\left\langle\sqrt{p_{i|B}}\right|$$

Probability sum of pure state density matrices
(with disjoint supports)
(de-lift of $\rho = \sum_k q_k \left|\psi_k\right\rangle\left\langle\psi_k\right|$)

- This is not an arbitrary probabilistic sum of pure state density matrices since the supports (the sets of points where the conditional distributions $\{p_{B,i}\}_{i\in U}$ are non-zero) are disjoint (which is the de-lift of dealing with the orthogonal representation of a quantum mixed state density operator).

Density matrix for a set-partition with probabilities: V

- Since the density matrix $\left|\sqrt{p_{i|B}}\right\rangle\left\langle\sqrt{p_{i|B}}\right|$ is multiplied by the probability p_B, it has the effect of canceling the denominator in the non-zero entries $\frac{\sqrt{p_i p_j}}{p_B}$ so that the whole mixed state density matrix p_π is a block-diagonal matrix.
- In the case of $n = 3$ and $\pi = \{\{0\}, \{1,2\}\}$ with point probabilities $\{p_i\}$, the mixed state density matrix p_π is:

$$p_\pi = p_0 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + (p_1 + p_2) \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{p_1}{p_1+p_2} & \frac{\sqrt{p_1 p_2}}{p_1+p_2} \\ 0 & \frac{\sqrt{p_1 p_2}}{p_1+p_2} & \frac{p_2}{p_1+p_2} \end{bmatrix}$$

$$= \begin{bmatrix} p_0 & 0 & 0 \\ 0 & p_1 & \sqrt{p_1 p_2} \\ 0 & \sqrt{p_1 p_2} & p_2 \end{bmatrix}.$$

Density matrix for a set-partition with probabilities: VI

- The blocks in the block-diagonal matrix p_π correspond to the blocks in the partition π (assuming a reordering of the indices so the indices in the same block are consecutive).
- Given a probability distribution $\{p_B\}_{B\in\pi}$ over the blocks in a partition, the logical entropy of the probability distribution is $h\left(\{p_B\}_{B\in\pi}\right) = 1 - \sum_{B\in\pi} p_B^2$.
- In the quantum case, the logical entropy of a mixed state density matrix ρ is $h(\rho) = 1 - \operatorname{tr}\left[\rho^2\right]$.
- Hence the de-lifted definition of the logical entropy of the set-version of the density matrix would be: $h(p_\pi) = 1 - \operatorname{tr}\left[p_\pi^2\right]$.

Density matrix for a set-partition with probabilities: VII

- The check that the de-lifting of the mixed state density matrix is correct is:

$$h\left(\{p_B\}_{B\in\pi}\right) = 1 - \sum_{B\in\pi} p_B^2 \stackrel{?}{=} 1 - \mathrm{tr}\left[p_\pi^2\right] = h\left(p_\pi\right).$$

- This is true in general but we can check it for the case at hand.

$$h\left(\{p_B\}_{B\in\pi}\right) = 1 - \sum_{B\in\pi} p_B^2 = 1 - p_0^2 - \left(p_1 + p_2\right)^2.$$

- The density matrix p_π squared is:

Density matrix for a set-partition with probabilities: VIII

$$p_\pi^2 = \begin{bmatrix} p_0^2 & 0 & 0 \\ 0 & p_1^2 + p_2 p_1 & (p_1 + p_2)\sqrt{p_1 p_2} \\ 0 & (p_1 + p_2)\sqrt{p_1 p_2} & p_2^2 + p_1 p_2 \end{bmatrix}$$

and

$$\begin{aligned} h(p_\pi) &= 1 - \operatorname{tr}\left[p_\pi^2\right] \\ &= 1 - \left[p_0^2 + p_1^2 + p_2 p_1 + p_2^2 + p_1 p_2\right] \\ &= 1 - p_0^2 - p_1(p_1 + p_2) - p_2(p_1 + p_2) \\ &= 1 - p_0^2 - (p_1 + p_2)^2 = h\left(\{p_B\}_{B \in \pi}\right) . \checkmark \end{aligned}$$

- Just as $p_{\mathbf{0}}$ is a pure state density matrix associated with the blob $\mathbf{0}$, the other pure state density matrices such as:

Density matrix for a set-partition with probabilities: IX

$$\left|\sqrt{p_{i|\{1,2\}}}\right\rangle\left\langle\sqrt{p_{i|\{1,2\}}}\right| = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{p_1}{p_1+p_2} & \frac{\sqrt{p_1 p_2}}{p_1+p_2} \\ 0 & \frac{\sqrt{p_1 p_2}}{p_1+p_2} & \frac{p_2}{p_1+p_2} \end{bmatrix}$$

represent mini-blobs or partition blocks where the non-zero support is a proper subset of U, e.g., in this case $B = \{1, 2\}$.

- The probability distribution is the conditional distribution $p_{i|\{1,2\}} = \Pr(i|\{1,2\})$ of $\{p_i\}$ conditioned on the event $\{1, 2\}$ which thus acts like a mini-blob.

Density matrix for a set-partition with probabilities: X

- This completes the density matrix treatment of set partitions π of U with a probability distribution $\{p_i\}_{i \in U}$ over the points of U which are the de-lifts of any density matrix $\rho = \sum_k q_k |\psi_k\rangle \langle\psi_k|$ represented in its orthogonal decomposition.

Special case of discrete partition: I

- Consider any general finite probability distribution $\{p_i\}_{i\in U}$.
- The point probabilities p_i are also the block probabilities for the discrete partition $\{\{i\}\}_{i\in U}$ on U which is usually denoted $\mathbf{1}$.
- In the case of $n = 3$, we have (since $p_i/p_B = p_i/p_i = 1$ if $B = \{i\}$):

$$p_{\mathbf{1}} = \sum_{\{i\}\in\mathbf{1}} p_i \left|\sqrt{p_{i|\{i\}}}\right\rangle \left\langle\sqrt{p_{i|\{i\}}}\right|$$

$$= p_0 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + p_1 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + p_2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} p_0 & 0 & 0 \\ 0 & p_1 & 0 \\ 0 & 0 & p_2 \end{bmatrix}.$$

Special case of discrete partition: II

- Squaring the matrix $p_{\mathbf{1}}$ just squares the diagonal terms so that:

$$h\left(p_{\mathbf{1}}\right) = 1 - \operatorname{tr}\left[p_{\mathbf{1}}^2\right] = 1 - \sum_{i \in U} p_i^2 = h\left(\{p_i\}\right).$$

Interpreting the density matrix entries: I

- We previously noted that a density matrix entry $\sqrt{p_i p_j}$ could be interpreted as the "amplitude" whose absolute square is the probability for the pair-outcome (i,j) in two independent draws from U according to the distribution $\{p_i\}_{i \in U}$.
- Now we have the more general case of a block-diagonal density matrix such as:

$$p_\pi = \begin{bmatrix} p_0 & 0 & 0 \\ 0 & p_1 & \sqrt{p_1 p_2} \\ 0 & \sqrt{p_1 p_2} & p_2 \end{bmatrix}.$$

Interpreting the density matrix entries: II

- Since the non-zero blocks correspond to the blocks $B \in \pi$, the non-zero entries $\sqrt{p_i p_j}$ correspond to pairs (i, j) where $i, j \in B$ for some $B \in \pi$, i.e., (i, j) is an indistinction or indit of the partition π.
- The zero terms in the block-diagonal matrix p_π are for the index pairs (i, j) where i and j are in different blocks of π, i.e., (i, j) is a distinction or dit of π.
- Hence for *all* the entries $(p_\pi)_{i,j}$, the absolute square is the probability that the pair (i, j) is drawn as an indistinction of π [where if (i, j) is a distinction of π, then the probability of it being an indistinction is 0].
- Thus we have the general interpretation of the p_π entries:

Interpreting the density matrix entries: III

$$(p_\pi)_{i,j} = \textit{indistinction amplitude} \text{ and}$$
$$(p_\pi)_{i,j}^2 = \textit{indistinction probability}.$$

- It is a general fact that:

$$\operatorname{tr}\left[p_\pi^2\right] = \sum_{i,j} (p_\pi)_{i,j}^2$$
$$= \text{sum of all indistinction probabilities.}$$

- The logical entropy is $h(p_\pi) = 1 - \operatorname{tr}\left[p_\pi^2\right]$ and $1 = \left(\sum_i p_i\right)^2 = \sum_{i,j} p_i p_j$, so we have:

$$h(p_\pi) = 1 - \operatorname{tr}\left[p_\pi^2\right] = \sum_{i,j}\left[p_i p_j - (p_\pi)_{i,j}^2\right].$$

Interpreting the density matrix entries: IV

- Now $p_i p_j$ is the probability of drawing the pair (i,j) in two independent draws from U according to the probabilities $\{p_i\}_{i\in U}$ regardless of whether (i,j) is a distinction or indistinction of π, and $(p_\pi)^2_{i,j}$ is the probability of drawing the pair (i,j) as an indistinction of π. Hence $p_i p_j - (p_\pi)^2_{i,j}$ is the probability of drawing the pair as a distinction (i,j) of π.
- Thus the interpretation of the $(p_\pi)^2_{i,j}$ terms as indistinction probabilities gives the interpretation of the logical entropy:

$$h(p_\pi) = \sum_{i,j}\left[p_i p_j - (p_\pi)^2_{i,j}\right]$$

= sum of all distinction probabilities.

- Hence the logical entropy of $h(p_\pi)$ is the probability that the two draws from U will give a distinction of π.

Change in entropy under refinement: I

- Suppose that $\pi \preceq \hat{\pi}$, i.e., $\hat{\pi}$ refines π as partitions of U (with the same point probabilities $\{p_i\}_{i \in U}$). For instance, $\hat{\pi}$ might be the discrete partition (like the result of a nondegenerate measurement).
  - $h(p_\pi) = \sum_{i,j} \left[p_i p_j - (p_\pi)_{i,j}^2 \right]$ (before refinement)
  - $h(p_{\hat{\pi}}) = \sum_{i,j} \left[p_i p_j - (p_{\hat{\pi}})_{i,j}^2 \right]$ (after refinement).
- When a partition π is refined to obtain a partition $\hat{\pi}$, certain pairs (i, j) which were indits of π become dits in $\hat{\pi}$. For such pairs (i, j), $(p_\pi)_{i,j}^2 = p_i p_j$ and $(p_{\hat{\pi}})_{i,j}^2 = 0$ and those are the *only differences* between the entropy formulas.

Change in entropy under refinement: II

- Hence the increase in entropy: $h(p_{\hat{\pi}}) - h(p_{\pi})$ is just the sum of the $p_i p_j$ terms, the pair-outcome probabilities, for the pairs that switch from being indits to being dits:

$$h(p_{\hat{\pi}}) - h(p_{\pi}) = \sum \left\{ p_i p_j \middle| (i,j) \in \operatorname{indit}(\pi) \cap \operatorname{dit}(\hat{\pi}) \right\}.$$

- Borrowing the language of coherence and decoherence from the quantum case, when i and j are in the same block of π, then $(p_{\pi})^2_{i,j} = p_i p_j$ and the density matrix entry $\sqrt{p_i p_j}$ is a *coherence term* since i and j *cohere* by being in the same block of π.
- But when the partition π is refined to get $\hat{\pi}$, if a π*-coherent* i and j are then in different blocks, then they have *decohered* and have been distinguished by $\hat{\pi}$.

Change in entropy under refinement: III

- The difference in the logical entropy:

$$h\left(p_{\hat{\pi}}\right) - h\left(p_{\pi}\right) = \sum \text{pair-probabilities } p_i p_j$$
$$\text{for the decohered pairs } (i, j).$$

- Incidentally, it might be noted how the interpretation of the density matrix entries is directly associated with the logical entropy. Matrix entries are about pairs (i, j) and so is the interpretation of logical entropy.

- For a quantum observable represented by the Hermitian operator T, then the average value of T in the state $\rho = \sum_k q_k |\psi_k\rangle \langle\psi_k|$ is:

$$\begin{aligned} \operatorname{tr}[T\rho] &= \sum_k q_k \operatorname{tr}[T |\psi_k\rangle \langle\psi_k|] \\ &= \sum_k q_k \langle\psi_k|T|\psi_k\rangle = \overline{\overline{T}} \end{aligned}$$

- The de-lift of a Hermitian operator T is a real-valued numerical attribute $T : U \to \mathbb{R}$ which, expressed as a matrix, is the diagonal matrix of its values $T(i)$ on the diagonal.
- The de-lift of ρ is: $p_\pi = \sum_k p_{B_k} |\sqrt{p_{i|B_k}}\rangle\langle\sqrt{p_{i|B_k}}|$.

- A diagonal matrix D times any matrix M has diagonal entries that are just the products $d_i m_i$ of the diagonal entries in the two matrices.
- Hence we have:

$$\mathrm{tr}\left[Tp_\pi\right] = \sum_k p_{B_k}\,\mathrm{tr}\left[T|\sqrt{p_{i|B_k}}\rangle\langle\sqrt{p_{i|B_k}}|\right]$$
$$= \sum_k p_{B_k}\sum_i T\left(i\right)p_i/p_{B_k} = \sum_i T\left(i\right)p_i = \overline{T}.$$

De-lifting table explained

Quantum case	Set case
$\rho = \sum_k q_k \left\|\psi_k\right\rangle \left\langle\psi_k\right\|$ (orth.de.)	$p_\pi = \sum_k p_{B_k} \left\|\sqrt{p_{i\|B_k}}\right\rangle \left\langle\sqrt{p_{i\|B_k}}\right\|$
Partition of eigenspaces $\{E_k\}$	Partition of blocks $\pi = \{B_k\}$
Orthonormal basis $\{\|i\rangle\}$	Set of points $U = \{i\}$
$\{p_i\}$ pdf associated with $\{\|i\rangle\}$	Pdf $\{p_i\}$ for $i \in U$
$\{q_k\}$ pdf associated with $\{E_k\}$	$\{p_B\}$ pdf associated with $\{B_k\}$
Pure states $\left\|\psi_k\right\rangle \left\langle\psi_k\right\|$	Pure states $\left\|\sqrt{p_{i\|B_k}}\right\rangle \left\langle\sqrt{p_{i\|B_k}}\right\|$
1-dimensional eigenspaces E_k	Discrete partition $\pi = 1$
Log. entropy $h(\rho) = 1 - \operatorname{tr}\left[\rho^2\right]$	Log. entropy $h(p_\pi) = 1 - \operatorname{tr}\left[p_\pi^2\right]$
$\left\|\rho_{ij}\right\|^2$ = indit probability	$(p_\pi)_{ij}^2$ = indit probability
$h(\rho) = \Sigma$ dit probabilities	$h(p_\pi) = \Sigma$ dit probabilities
$h(\hat{\rho}) - h(\rho) = \Sigma$ new dit probs.	$h(p_{\hat{\pi}}) - h(p_\pi) = \Sigma$ new dit probs
$\overline{T} = \operatorname{tr}[T\rho]$	$\overline{T} = \operatorname{tr}[Tp_\pi]$

Products of sets and vector spaces

–And how to interpret reduced density matrices

David Ellerman

UCR

May 2012

Lifting Direct to Tensor Products

- One yoga of lifting says: "Apply the set-concept to basis sets to generate the vector-space concept."
- Given finite sets X and Y, the *direct product* $X \times Y$ is the set of all ordered pairs (x, y). Given a basis $\{|i\rangle\}$ for the vector space H^A and a basis $\{|j\rangle\}$ for H^B (both finite dimensional), the direct product of the basis sets gives $\{|i\rangle \otimes |j\rangle\}$ which is a basis for the *tensor product* (NB: not the direct product) of vector spaces $H^A \otimes H^B$, so tensor products of vector spaces are the lift of direct products of sets.
- Since we can de-lift density matrices, we can de-lift the density matrix treatment of states ρ^{AB} on $H^A \otimes H^B$ to sets.
- One goal is to make sense out of the problem that the ignorance interpretation of density matrices does not apply to the reduced density matrices ρ^A and ρ^B for a pure entangled state ρ^{AB} on $H^A \otimes H^B$.

Density matrices for joint distributions: I

- Given finite sets $X = \{0, 1, ..., m-1\}$ and $Y = \{0, 1, ..., n-1\}$ and a joint probability distribution $p(x, y) = p_{xy}$ on the direct product $X \times Y$. Taking the indiscrete partition $\mathbf{0}$ as the partition on $X \times Y$, then $mn \times mn$ density matrix is:

$$p_{\mathbf{0}} = \begin{bmatrix} \sqrt{p_{00}} \\ \sqrt{p_{01}} \\ \vdots \\ \sqrt{p_{11}} \end{bmatrix} \begin{bmatrix} \sqrt{p_{00}} & \sqrt{p_{01}} & \cdots & \sqrt{p_{11}} \end{bmatrix}$$

$$= \begin{bmatrix} p_{00} & \sqrt{p_{00}p_{01}} & \cdots & \sqrt{p_{00}p_{m-1,n-1}} \\ \sqrt{p_{00}p_{01}} & p_{01} & \cdots & \sqrt{p_{01}p_{m-1,n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{p_{00}p_{m-1,n-1}} & \sqrt{p_{01}p_{m-1,n-1}} & \cdots & p_{m-1,n-1} \end{bmatrix}.$$

Density matrices for joint distributions: II

- For a matrix derivation of the marginal distributions $p_x = \sum_y p_{xy}$ and $p_y = \sum_x p_{xy}$, arrange the coefficients $\sqrt{p_{ij}}$ in an $m \times n$ matrix $[\alpha] = \left[\sqrt{p_{ij}}\right]$, and then $[\alpha]\,[\alpha]^t$ is the symmetric, positive, and unit trace $m \times m$ *density matrix for the marginal distribution* p_x, e.g., for $m, n = 3$,

$$[\alpha]\,[\alpha]^t =$$

$$\begin{bmatrix} \sqrt{p_{00}} & \sqrt{p_{01}} & \sqrt{p_{0,2}} \\ \sqrt{p_{10}} & \sqrt{p_{11}} & \sqrt{p_{1,2}} \\ \sqrt{p_{20}} & \sqrt{p_{21}} & \sqrt{p_{22}} \end{bmatrix} \begin{bmatrix} \sqrt{p_{00}} & \sqrt{p_{10}} & \sqrt{p_{20}} \\ \sqrt{p_{01}} & \sqrt{p_{11}} & \sqrt{p_{21}} \\ \sqrt{p_{02}} & \sqrt{p_{12}} & \sqrt{p_{22}} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_y p_{0y} & \sum_y \sqrt{p_{0y}p_{1y}} & \sum_y \sqrt{p_{0y}p_{2y}} \\ \sum_y \sqrt{p_{1y}p_{0y}} & \sum_y p_{1y} & \sum_y \sqrt{p_{1y}p_{2y}} \\ \sum_y \sqrt{p_{2y}p_{0y}} & \sum_y \sqrt{p_{2y}p_{1y}} & \sum_y p_{2y} \end{bmatrix}.$$

Density matrices for joint distributions: III

- Note that this is not a pure state density matrix since, in general, the off-diagonal entries are not the square root of the product of the diagonal entries:

$$\sum_y \sqrt{p_{iy}p_{ky}} \neq \sqrt{\sum_y p_{iy}}\sqrt{\sum_y p_{ky}} = \sqrt{p_i p_k} \text{ where } i,k \in X.$$

- But p_{xy} is an independent joint distribution if $\forall x, y$, $p_{xy} = p_x p_y$. Then and only then the off-diagonal terms in the marginal distribution matrix are:

$$\sum_y \sqrt{p_{iy}p_{ky}} = \sum_y \sqrt{p_i p_y p_k p_y} = \sum_y p_y \sqrt{p_i p_k} = \sqrt{p_i p_k}, \text{ i.e.,}$$

the marginal distribution matrix is a pure state matrix
iff $p_{xy} = p_x p_y \ \forall x, y$ (i.e., p_{xy} independent)

Density matrices for joint distributions: IV

- All of this carries over, *mutatus mutandis*, to the quantum case where the reduced density matrices ρ^A and ρ^B play the role of the marginal distributions p_x and p_y:
  - $\rho^{AB} = |\psi\rangle \langle\psi|$ where $|\psi\rangle = \sum_{i,j} \alpha_{ij} |i\rangle \otimes |j\rangle$,
  - $\rho^A = [\alpha] [\alpha]^\dagger$ and $\rho^B = [\alpha]^\dagger [\alpha]$, and
  - ρ^A and ρ^B are pure states iff $\rho^{AB} = \rho^A \otimes \rho^B$.
- Restricting attention to the set analogue to the two qubit case, we have the pure state density matrix:

Density matrices for joint distributions: V

$$p_{\mathbf{0}} = \begin{bmatrix} \sqrt{p_{00}} \\ \sqrt{p_{01}} \\ \sqrt{p_{10}} \\ \sqrt{p_{11}} \end{bmatrix} \begin{bmatrix} \sqrt{p_{00}} & \sqrt{p_{01}} & \sqrt{p_{10}} & \sqrt{p_{11}} \end{bmatrix}$$

$$= \begin{bmatrix} p_{00} & \sqrt{p_{00}p_{01}} & \sqrt{p_{00}p_{10}} & \sqrt{p_{00}p_{11}} \\ \sqrt{p_{00}p_{01}} & p_{01} & \sqrt{p_{01}p_{10}} & \sqrt{p_{01}p_{11}} \\ \sqrt{p_{00}p_{10}} & \sqrt{p_{01}p_{10}} & p_{10} & \sqrt{p_{10}p_{11}} \\ \sqrt{p_{00}p_{11}} & \sqrt{p_{01}p_{11}} & \sqrt{p_{10}p_{11}} & p_{11} \end{bmatrix}.$$

- The "alpha-coefficient matrix" is $[\alpha] = \begin{bmatrix} \sqrt{p_{00}} & \sqrt{p_{01}} \\ \sqrt{p_{10}} & \sqrt{p_{11}} \end{bmatrix}$ and

$$p_x = [\alpha] [\alpha]^t = \begin{bmatrix} p_{0_X} & \sqrt{p_{00}p_{10}} + \sqrt{p_{01}p_{11}} \\ \sqrt{p_{00}p_{10}} + \sqrt{p_{01}p_{11}} & p_{1_X} \end{bmatrix}$$

Density matrices for joint distributions: VI

and

$$p_y = [\alpha]^t [\alpha] = \begin{bmatrix} p_{0_Y} & \sqrt{p_{00}p_{01}} + \sqrt{p_{10}p_{11}} \\ \sqrt{p_{00}p_{01}} + \sqrt{p_{10}p_{11}} & p_{1_Y} \end{bmatrix}.$$

- A *correlated* (i.e., non-independent) joint distribution p_{xy} on $X \times Y$ (using the set partition **0**) is the set analogue of an entangled (pure) state ρ^{AB} on $H^A \otimes H^B$ in the quantum case so:

Correlated p_{xy} lifts to entangled ρ^{AB}
(and entangled ρ^{AB} de-lifts to correlated p_{xy}).

Density matrices for joint distributions: VII

- Given a numerical attribute $T : X \rightarrow \mathbb{R}$ on X, the marginal distribution p_x has all the information needed to determine the average value of T according to the joint distribution p_{xy} since:

$$\overline{T} = \sum_{x,y} p_{xy} T(x) = \sum_x T(x) \sum_y p_{xy} = \sum_x p_x T(x).$$

- The lift of an attribute $T : X \rightarrow \mathbb{R}$ is a Hermitian operator $\hat{T}$ on H^A, the lift of the marginal distribution p_x is the reduced density matrix ρ^A, and the lift for the average value is:

$$\langle \rho^{AB} | \hat{T} \otimes I | \rho^{AB} \rangle = \operatorname{tr}\left[\left(\hat{T} \otimes I \right) \rho^{AB} \right] = \operatorname{tr}\left[\hat{T} \rho^A \right] = \langle \rho^A | \hat{T} | \rho^A \rangle.$$

Interpreting the reduced density matrices: I

- In the set case, the reduced density matrices $p_x = [\alpha]\,[\alpha]^t$ and $p_y = [\alpha]^t\,[\alpha]$ for the marginal distributions of an "entangled" joint distribution p_{xy} are something new. They are *not* density matrices p_π associated with a point distribution $\{p_i\}$ and a partition π on the set of points.
- Previously, for our set versions of the density matrices p_π, the off-diagonal terms were either
  - $\sqrt{p_i p_k}$ (where p_i and p_k were the corresponding diagonal terms) when (i,k) was an indistinction of the set partition π (i.e., were in the same block); or
  - 0 when (i,k) was a distinction of the set partition π (i.e., were in different blocks).

Interpreting the reduced density matrices: II

- In either case, the off-diagonal term was an "indistinction amplitude" whose absolute square was the indistinction (according to π) probability in a pair of independent draws from the distribution $\{p_i\}$.
- This 0 or $\sqrt{p_i p_k}$ nature of the off-diagonal terms resulted from de-lifting the orthogonal decomposition of a density matrix ρ so that we had disjoint blocks B in the partition on U. Now we are forming density matrices ρ^A and ρ^B. Later we will give a canonical decomposition of the reduced density matrices into a probability-weighted sum of pure states (not necessarily orthogonal).

Interpreting the reduced density matrices: III

- For the reduced density matrices p_x and p_y of an entangled p_{xy}, the off-diagonal terms (intermediate between the extremes $\sqrt{p_i p_k}$ and 0) still represent indistinction amplitudes but a pair (i,k) can be partially indistinct and partially distinct (instead of the previous all-or-nothing indistinction)!
- That is, instead of being identified or not by a partition π, a pair of x's (i,k) can partially overlap (where no overlap means "distinction" and total overlap means "indistinction").
- Focusing on p_x for illustrative purposes, a pair (i,k) for $i,k \in X$ overlap or are "indistinct" insofar as they are linked by the y-probabilities p_{iy} and p_{ky}.

Interpreting the reduced density matrices: IV

- Hence the off-diagonal term $\sum_y \sqrt{p_{iy}p_{ky}}$ is the sum of the indistinction or overlap amplitudes for the pair (i,k), and the (absolute) square $\left(\sum_y \sqrt{p_{iy}p_{ky}}\right)^2$ is the *indistinction or overlap probability* for (i,k).
- The sum of those absolute squares is the trace of the square of the density matrix p_x and their complement is the logical entropy of p_x.
- Since $\sum_{i,k} p_i p_k = 1$, the logical entropy is:

$$h(p_x) = 1 - \operatorname{tr}\left[p_x^2\right] = \sum_{i,k}\left[p_i p_k - \left(\sum_y \sqrt{p_{iy}p_{ky}}\right)^2\right].$$

Interpreting the reduced density matrices: V

- Since $p_i p_k$ is the probability of getting the pair (i,k) in a pair of draws from the marginal distribution p_x and $\left(\sum_y \sqrt{p_{iy}p_{ky}}\right)^2$ is the indistinction probability for (i,k), the difference is the distinction probability. Hence we have:

$$h(p_x) = \sum_{i,k}\left[p_i p_k - \left(\sum_y \sqrt{p_{iy}p_{ky}}\right)^2\right]$$

Logical entropy = $\sum$ [distinction probabilities].

- Although we have a more general structure than a set partition with point probabilities (due to the partial overlaps), the idea of logical entropy as the sum of distinction probabilities still survives.
- All this lifts, *mutatis mutandis*, to the quantum case.

Bijective or "Schmidt" special case: I

- Suppose the support of p_{xy} on $X \times Y$ is the graph of a bijection $X \leftrightarrow Y$ (the images in X and Y have the same cardinality n) so that $p_{x_i y_i} = p(x_i, y_i) = p_i$ for $i = 0, 1, ..., n-1$ and $p_{xy} = 0$ otherwise.
- The probabilities p_i are the set versions of the Schmidt coefficients $\sqrt{p_i}$ squared in the quantum case.
- In the set version of the two qubit model, suppose the bijection is $0_X \leftrightarrow 0_Y$ and $1_X \leftrightarrow 1_Y$ so the only two non-zero probabilities are p_{00} and p_{11} on the mini-blob subset $B = \{(0,0), (1,1)\} \subseteq X \times Y$. Then the density matrix is:

Bijective or "Schmidt" special case: II

$$\left|\sqrt{p_{i|\{(0,0),(1,1)\}}}\right\rangle\left\langle\sqrt{p_{i|\{(0,0),(1,1)\}}}\right| = \begin{bmatrix}\sqrt{p_{00}}\\0\\0\\\sqrt{p_{11}}\end{bmatrix}\begin{bmatrix}\sqrt{p_{00}} & 0 & 0 & \sqrt{p_{11}}\end{bmatrix}$$

$$= \begin{bmatrix} p_{00} & 0 & 0 & \sqrt{p_{00}p_{11}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \sqrt{p_{00}p_{11}} & 0 & 0 & p_{11}\end{bmatrix}.$$

- The matrix of coefficients is $[\alpha] = \begin{bmatrix}\sqrt{p_{00}} & 0 \\ 0 & \sqrt{p_{11}}\end{bmatrix}$ so the marginal distributions are:

$$p_x = [\alpha][\alpha]^t = \begin{bmatrix} p_{00} & 0 \\ 0 & p_{11}\end{bmatrix} = [\alpha]^t[\alpha] = p_y.$$

Bijective or "Schmidt" special case: III

- Since the support of p_{xy} is a bijection, there are no y's connecting two different x's (and vice-versa) so there are no partial overlaps in the density matrices for the marginal distributions. They are the density matrices for the point distribution $\{p_{00}, p_{11}\}$ and the discrete partition on X and on Y.

Ignorance interpretation does not apply?: I

- In the quantum case, a density matrix ρ has the general form as a probabilistic mixture of pure states:

$$\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|.$$

- The usual interpretation is as a statistical ensemble consisting of the pure state $\rho_i = \left|\psi_i\right\rangle \left\langle\psi_i\right|$ with probability p_i, which is sometimes called the *ignorance interpretation*.
- But when the density matrix is a reduced density matrix ρ^A or ρ^B from an entangled pure state ρ^{AB} on $H^A \otimes H^B$, then it is said that "the ignorance interpretation does not apply."
- It is said that "the ignorance interpretation does not apply" since applying it to $\rho^A = \sum_i p_i \rho_i^A$ and to $\rho^B = \sum_j q_j \rho_j^B$, we have:

Ignorance interpretation does not apply?: II

- the component system A is in a pure state ρ_i^A with probability p_i, and
- the component system B is in a pure state ρ_j^B with probability q_j, so
- the joint system AB represented in $H^A \otimes H^B$ is in the pure product state $\rho_i^A \otimes \rho_j^B$ with probability $p_i q_j$.
- Thus the composite system is in the mixed state $\sum_{i,j} p_i q_j \rho_i^A \otimes \rho_j^B$ contrary to it being in the entangled pure state ρ^{AB}.
- Hence "the ignorance interpretation does not apply" to the reduced density matrices of pure entangled states. Yet they are certainly density matrices, so it is said that "there are two kinds of density matrices."

- For instance, Paul Busch talks about "two fundamentally different uses of mixed state ('density') operators":

Ignorance interpretation does not apply?: III

"It is important to note that a pure entangled state of a compound system necessarily yields a mixed state description for each of its subsystems, and that these density operators of the subsystems do not allow an ignorance interpretation." [Busch, P. 2002. Classical versus quantum ontology. Studies in History and Philosophy of Modern Physics. 33: 517-539, p. 526]

- Or Bernard D'Espagnat:

Ignorance interpretation does not apply?: IV

"Since mixtures of the E type [i.e., ρ^A or ρ^B] and mixtures of the Ê type [i.e., ordinary ρ] are in principle operationally different concepts, it is appropriate, at least when fundamental problems are discussed, to differentiate them... . In what follows, the expressions proper and improper mixtures are used to designate mixtures of the Ê and E types, respectively." [D'Espagnat, Bernard 1999. Conceptual Foundations of Quantum Mechanics (2nd ed.), Reading MA: Perseus Books. p. 61]

Canonical decomposition of reduced density matrices: I

- In the set case, probabilities occur as sampling probabilities for outcomes in the sample space.
- In QM, probabilities occur in two ways:
  - the *sampling probabilities* q_k involved in a statistical ensemble $\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|$ which are used in the ignorance interpretation and
  - the *measurement probabilities* for the outcome of a measurement.

Canonical decomposition of reduced density matrices: II

- In the reduced density matrix $\rho^B = [\alpha]^\dagger [\alpha]$ of a pure state ρ^{AB}, the probabilities on the diagonal, $p_j = \sum_i \alpha_{ij}\alpha_{ij}^*$, are *measurement* probabilities, e.g., $p_j = \sum_i \alpha_{ij}\alpha_{ij}^*$ is the probability of getting the outcome $|j\rangle$ in a measurement of the second component of the state $\rho = |\psi\rangle\langle\psi|$ where $|\psi\rangle = \sum_{i,j} \alpha_{ij} |i\rangle \otimes |j\rangle$ (using the measurement basis $\{|j\rangle\}$).
- We will present a canonical decomposition of a reduced density matrix as a probabilistic sum of pure states where the sampling probabilities over the pure states are those measurement probabilities.
- The set case is presented with the quantum case an easy generalization (*mutatis mutandis*).

Canonical decomposition of reduced density matrices: III

- The setting is a joint probability distribution $p_{XY}(x,y) = p_{xy}$ on the direct product $X \times Y$ of two finite samples spaces. To construct the reduced density matrix for p_X, the marginal distribution over X, we construct the pure density matrix for the conditional probability distribution $\Pr(x|y)$ over X where the value of y is given. The sum of these pure density matrices weighted by the probabilities for getting the values of y, i.e., $p_Y(y) = \sum_x p_{xy}$, gives the canonical decomposition for the reduced density matrix for p_X. Reversing the role of X and Y gives the canonical decomposition for the reduced density matrix p_Y.

Canonical decomposition of reduced density matrices: IV

- For illustrative purposes, take $X = \{0, 1, 2\}$ and $Y = \{0, 1\}$. Then the pure density matrix for the conditional distribution $\Pr(x|y)$ is:

$$\begin{bmatrix} p_{0y}/p_Y(y) & \sqrt{p_{0y}p_{1y}}/p_Y(y) & \sqrt{p_{0y}p_{2y}}/p_Y(y) \\ \sqrt{p_{0y}p_{1y}}/p_Y(y) & p_{1y}/p_Y(y) & \sqrt{p_{1y}p_{2y}}/p_Y(y) \\ \sqrt{p_{0y}p_{2y}}/p_Y(y) & \sqrt{p_{1y}p_{2y}}/p_Y(y) & p_{2y}/p_Y(y) \end{bmatrix}$$

and the probability-weighted sum of these pure density matrices is the reduced density matrix for the marginal distribution over X:

Canonical decomposition of reduced density matrices: V

$$p_X = \begin{bmatrix} \sum_y p_{0y} & \sum_y \sqrt{p_{0y}p_{1y}} & \sum_y \sqrt{p_{0y}p_{2y}} \\ \sum_y \sqrt{p_{1y}p_{0y}} & \sum_y p_{1y} & \sum_y \sqrt{p_{1y}p_{2y}} \\ \sum_y \sqrt{p_{2y}p_{0y}} & \sum_y \sqrt{p_{2y}p_{1y}} & \sum_y p_{2y} \end{bmatrix} =$$

$$\sum_y p_Y(y) \begin{bmatrix} p_{0y}/p_Y(y) & \sqrt{p_{0y}p_{1y}}/p_Y(y) & \sqrt{p_{0y}p_{2y}}/p_Y(y) \\ \sqrt{p_{0y}p_{1y}}/p_Y(y) & p_{1y}/p_Y(y) & \sqrt{p_{1y}p_{2y}}/p_Y(y) \\ \sqrt{p_{0y}p_{2y}}/p_Y(y) & \sqrt{p_{1y}p_{2y}}/p_Y(y) & p_{2y}/p_Y(y) \end{bmatrix}.$$

Canonical decomposition of reduced density matrix p_X

- The usual definition of the single probability number in the marginal distribution $p_X(x)$ is

$$p_X(x) = \sum_y p_{XY}(x,y) = \sum_y p_Y(y) \Pr(x|y).$$

Canonical decomposition of reduced density matrices: VI

- The above canonical expression for the reduced density matrix for p_X is just the density matrix version of that summation. The probabilities $p_Y(y)$ in the sum of the weighted pure state density matrices are the measurement probabilities for measuring or sampling the second component.
- In each pure state matrix for $\Pr(x|y)$, the off-diagonal terms $\sqrt{p_{iy}p_{jy}}/p_Y(y)$ give the indistinction amplitudes for an X-pair (i,j) where $i,j \in X$ and the square is the indistinction probability for the X-pair given that y-value.
- If the original joint distribution is independent, i.e., $p_{XY}(x,y) = p_X(x)\,p_Y(y)$, then all the pure state density matrices for the conditional distributions are the same since:

Canonical decomposition of reduced density matrices: VII

$$\sqrt{p_{iy}p_{jy}}/p_Y(y) = \sqrt{p_X(i)\,p_Y(y)\,p_X(j)\,p_Y(y)}/p_Y(y) = \sqrt{p_X(i)\,p_X(j)}$$

independent of y, and all the probability weights $p_Y(y)$ sum to 1 so the reduced density matrix for X has the entries: $\sqrt{p_X(i)\,p_X(j)}$ and similarly for Y. And the Hadamard product of those reduced density matrices gives the density matrix for the independent joint distribution.

- The opposite to having all the pure state density matrices $\Pr(x|y)$ being the same is to have them all be orthogonal, and that occurs in the opposite situation of "maximal entanglement" where the support of p_{XY} is a bijection.
- These results extend, *mutatis mutandis*, to the quantum case.

Bijective set example: I

- For instance, we can use the bijective special case considered above with the pure state distribution with probability p_{00} for $(0_X, 0_Y)$ and p_{11} for $(1_X, 1_Y)$ where $|X| = 2 = |Y|$.
- Then $p_Y(0) = p_{00}$ and $p_Y(1) = p_{11}$, and the two pure state density matrices are the two orthogonal matrices:

$$\Pr(x|0) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } \Pr(x|1) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

and the reduced density matrix for X is:

$$p_{00} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + p_{11} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} p_{00} & 0 \\ 0 & p_{11} \end{bmatrix} = p_X.$$

Bijective set example: II

- When we apply independent sampling to the marginal distributions as stand alone pdfs, then we obtain, say, a draw of 0_Y from Y with probability p_{00} and a draw of 1_X from X with probability p_{11}, which together means a draw of $(1_X, 0_Y)$ from the "compound urn" $X \times Y$ with probability $p_{00}p_{11}$.
- But the probability of $(1_X, 0_Y)$ is $p_{10} = 0$ so something has gone wrong.
- The problem lies in the implicit assumption that one can *independently* draw from the Y-urn and from the X-urn and get the probabilities for the pair-outcomes (x, y) when the two urns are "entangled" by the joint distribution p_{XY}.

Bijective set example: III

- The canonical presentation of the reduced density matrices automatically takes the entanglement into account when one applies the "ignorance interpretation" using that presentation–since the probability coefficients for say p_X are the Y-measurement probabilities for the other component.
- Thus to compute the probability of $(1_X, 0_Y)$, we first measure Y and get 0_Y with probability p_{00} which corresponds, using the ignorance interpretation, to taking the pure state

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ in the canonical } p_X = p_{00} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + p_{11} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

and then we see the probability of 1_X is 0 in that pure state.

- Thus applying the "ignorance interpretation" to the canonical representation of the reduced density matrices will give the correct results for the sampling/measurements of the joint distribution.

Reflections on mixed states and ignorance interpretation: I

- The "ignorance interpretation" of a mixed state $\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|$ in terms of the "state" resulting from a physicist sampling a probability distribution $\{p_i\}$ to decide which pure state $|\psi_i\rangle$ to prepare–was always a bit of a fantasy.
- The point is that Nature, in effect, samples a probability distribution $\{\alpha_i \alpha_i^*\}$ to decide which pure state $|i\rangle$ will result from a measurement of $|\psi\rangle = \sum_i \alpha_i |i\rangle$ so the general result of the measurement can be described by the mixed state: $\rho = \sum_i \alpha_i \alpha_i^* |i\rangle \langle i|$.

Reflections on mixed states and ignorance interpretation: II

- And then that general interpretation got into trouble when applied to the reduced density matrices ρ^A and ρ^B resulting from an entangled ρ^{AB}.
  1. The first thing that got the ignorance interpretation into trouble for ρ^A and ρ^B was the assumption that they are the "states" of the component systems when ρ^{AB} is the state of the composite system. Instead they describe the results of certain measurements.
  2. The second point is that the two component systems cannot be measured or sampled independently when the composite state is entangled.

Reflections on mixed states and ignorance interpretation: III

- We have given a canonical representation of the mixed state reduced density matrices–as a probabilistic mixture of pure state density matrices–so that the "ignorance interpretation" of either reduced density matrix involves essentially the measurements on the joint system that takes into account the entanglement.
- In this manner, the "ignorance interpretation" can be applied to reduced density matrices and it yields a correct account of the measurement results of the joint system.

Reflections on mixed states and ignorance interpretation: IV

- For instance, in the set case, to compute the probability of the result (x, y) using the reduced density matrix for $p_X = \sum_y p_Y(y) \Pr(x|y)$, we in effect sample to get y with the probability $p_Y(y)$ which means taking the pure density matrix $\Pr(x|y)$ with the diagonal probabilities $p_{xy}/p_Y(y)$ for the outcome x so the probability of getting (x, y), i.e., x by way of y, is the correct:

$$p_Y(y) \frac{p_{XY}(x,y)}{p_Y(y)} = p_{XY}(x, y).$$

Products summary

Set case	Quantum case
Direct product $X \times Y$	Tensor product $H^A \otimes H^B$
$\{p_{xy}\}$ joint pdf on $X \times Y$	Pure ρ^{AB} on $H^A \otimes H^B$
$\{p_x\}$, $\{p_y\}$ marginals	ρ^A, ρ^B reduced density ops
$p_{xy} = p_x p_y$ indep. dist.	$\rho^{AB} = \rho^A \otimes \rho^B$ product state
$p_{xy} \neq p_x p_y$ correlated	$\rho^{AB} \neq \rho^A \otimes \rho^B$ entangled state
p_{xy} indep. $\Rightarrow p_x, p_y$ pure	Product $\rho^{AB} \Rightarrow \rho^A, \rho^B$ pure
Correlated $p_{xy} \Rightarrow p_x, p_y$ mixed	Entangled $\rho^{AB} \Rightarrow \rho^A, \rho^B$ mixed
$\sum_{x,y} p_{xy} T(x) = \sum_x p_x T(x)$	$\langle \rho^{AB} \vert \hat{T} \otimes I \vert \rho^{AB} \rangle = \langle \rho^A \vert \hat{T} \vert \rho^A \rangle$
Apply ig. int. to canonical rep	Apply ig. int. to canonical rep

Shannon Noiseless Coding Theorem

David Ellerman

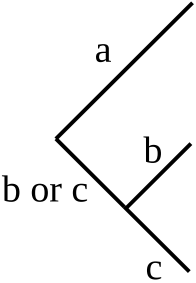
UCR

May 2012

Simple coding examples: I

- General idea: to find the average minimum number of equiprobable binary questions needed per letter to identity a probabilistic message. That is the interpretation of the Shannon entropy $H(p)$.
- Suppose the message is one letter, a, b, or c, and that $p_a = \frac{1}{2}$ while $p_b = \frac{1}{4} = p_c$.
- Then the picture is:

Simple coding examples: II



- The pattern of answers to the binary questions give a binary code for the messages where:
  - $''a'' = 1;$
  - $''b'' = 01;$
  - $''c'' = 00.$

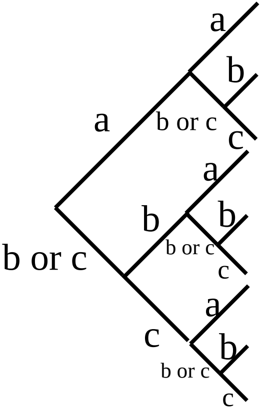
Simple coding examples: III

- Thus the average #binary-coded questions, where $\#(01) = 2$ is the number of binary digits in the code, is:

$$\frac{1}{2}\#(1) + \frac{1}{4}\#(01) + \frac{1}{4}\#(00) = \frac{1}{2} + \frac{2}{4} + \frac{2}{4} = \frac{3}{2}$$
$$= H(p) = \sum_i p_i \log_2\left(\frac{1}{p_i}\right) = \frac{1}{2}\log(2) + \frac{1}{4}\log(4) + \frac{1}{4}\log(4).$$

- Now consider $N = 2$ letter messages with the same probability distribution. Then the picture with the binary codes is:

Simple coding examples: IV



aa = 11 $\#(11)/2^2$

ab = 101 $\#(101)/2^3$

ac = 100 $\#(100)/2^3$

ba = 011 $\#(011)/2^3$

bb = 0101 $\#(0101)/2^4$

bc = 0100 $\#(0100)/2^4$

ca = 001 $\#(001)/2^3$

cb = 0001 $\#(0001)/2^4$

cc = 0000 $\#(0000)/2^4$

Simple coding examples: V

- Again the average number of questions necessary to identify the message is:

$$\frac{2}{4}+\frac{3}{8}+\frac{3}{8}+\frac{3}{8}+\frac{4}{16}+\frac{4}{16}+\frac{3}{8}+\frac{4}{16}+\frac{4}{16}$$
$$=\frac{48}{16}=3=2H\left(p\right)$$

- And thus the average number of questions needed *per letter* in the message is $NH\left(p\right)/N=H\left(p\right)$.
- These examples are somewhat artificial since the probabilities are all (negative) powers of 2 so we immediately have a most efficient questioning scheme to find the message and thus we have the minimum average number of equiprobable binary questions ("bits") needed per letter to find the message.

A statistical example: I

- Suppose $p_a = p_b = p_c = \frac{1}{3}$. Then a one-letter or two-letter message cannot be exactly coded with a binary $0, 1$ code with equiprobable 0's and 1's.
- But any probability can be better and better approximated by longer and longer representations in the binary number system.
- Hence we can consider longer and longer messages of N letters along with better and better approximations with binary codes.
- The long run behavior of messages $u_1 u_2 ... u_N$ where $u_i \in \{a, b, c\}$ is modeled by the law of large numbers so that the letter a will tend to occur $p_a N = \frac{1}{3} N$ times and similarly for b and c.

A statistical example: II

- Such a message is called *typical*.
- The probability of any one of those typical messages is:

$$p_a^{p_a N} p_b^{p_b N} p_c^{p_c N} = \left[p_a^{p_a} p_b^{p_b} p_c^{p_c} \right]^N$$

or, in this case,

$$\left[\left(\frac{1}{3}\right)^{1/3} \left(\frac{1}{3}\right)^{1/3} \left(\frac{1}{3}\right)^{1/3} \right]^N = \left(\frac{1}{3}\right)^N.$$

- Hence the number of such typical messages is 3^N.

A statistical example: III

- If each message was assigned a unique binary code, then the number of $0,1$'s in the code would have to be X where $2^X = 3^N$ or $X = \log_2\left(3^N\right) = N\log_2(3)$. Hence the number of equiprobable binary questions or bits needed per letter of the messages is:

$$N\log_2(3)/N = \log_2(3) = 3 \times \frac{1}{3}\log_2\left(\frac{1}{1/3}\right) = H(p).$$

- This example shows the general pattern.

Shannon Noiseless Coding Theorem: I

- Let $p = (p_1, ..., p_n)$ be the probabilities over a n-letter alphabet $A = \{a_1, ..., a_n\}$.
- In an N-letter message, the probability of a particular message $u_1 u_2 ... u_N$ is $\Pi_{i=1}^{N} \Pr(u_i)$ where u_i could be any of the symbols in the alphabet.
- In a typical message, the i^{th} symbol will occur $p_i N$ times so the probability of a typical message is (note change of indices):

$$\Pi_{k=1}^{n} p_k^{p_k N} = \left[\Pi_{k=1}^{n} p_k^{p_k}\right]^N .$$

Shannon Noiseless Coding Theorem: II

- Since the typical messages are equiprobable, the number of typical messages is $\left[\Pi_{k=1}^{n} p_k^{-p_k}\right]^N$ and assigning a unique binary code to each typical message requires X bits where $2^X = \left[\Pi_{k=1}^{n} p_k^{-p_k}\right]^N$ so that:

$$X = \log_2 \left\{ \left[\Pi_{k=1}^{n} p_k^{-p_k}\right]^N \right\} = N \log_2 \left[\Pi_{k=1}^{n} p_k^{-p_k}\right]$$
$$= N \sum_k \log_2 \left(p_k^{-p_k}\right) = N \sum_k -p_k \log_2 (p_k)$$
$$= N \sum_k p_k \log_2 \left(\frac{1}{p_k}\right) = NH(p).$$

Shannon Noiseless Coding Theorem: III

- Hence the Shannon entropy $H(p) = \sum_k p_k \log_2 \left(\frac{1}{p_k}\right)$ is interpreted as the average number of bits necessary per letter in the message.
- It is in this context of coding and communication that Shannon's *Mathematical Theory of Communication* supplies the appropriate concept, not in the foundations of information theory.

Schmacher's quantum version: I

- Alice is sending a message to Bob. Instead of seeing the message being generated by a probability distribution $p = \{p_1, p_2, ..., p_n\}$ over a set of n letters $A = \{a_1, ..., a_n\}$, we think of a mixed state $\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|$.
- But since the states $|\psi_i\rangle$, need not be orthogonal, we move to the orthogonal decomposition of the mixed state: $\rho = \sum_i q_i |\varphi_i\rangle \langle\varphi_i|$ where $\left\langle \varphi_i | \varphi_j \right\rangle = \delta_{ij}$.
- Then we can apply the classical Shannon theorem to a classical typical sequence of N states $|\varphi_i\rangle$ with the probabilities $q = \{q_i\}$. The sequences are represented in ever larger Hilbert spaces $H^{\otimes N}$ and the typical sequences span the *typical subspace* at each stage.

Schmacher's quantum version: II

- The limit on "compression" is again given by $NH(q) = NS(\rho)$ so the von Neumann entropy $S(\rho)$ gives the number of qubits needed per state $|\varphi_i\rangle$ to code a typical sequence.
- Needless to say, this is essentially a straight lifting of Shannon's theorem to the quantum case:
  - The set of n letters $A = \{a_1, ..., a_n\}$ are all completely distinct from each other, and that lifts to the orthogonal states $\{|\varphi_i\rangle\}$ in the orthogonal decomposition of ρ;
  - A sequence of N letters from $A = \{a_1, ..., a_n\}$ is an element of the direct product A^N lifts to a "sequence" of states $\left|\varphi_{i_1}\right\rangle \otimes \left|\varphi_{i_2}\right\rangle \otimes ... \otimes \left|\varphi_{i_N}\right\rangle$ which is an element of the tensor product $H^{\otimes N}$;

Schmacher's quantum version: III

- The subset of typical sequences of each power A^N lifts to the subspace generated by the typical sequences in $H^{\otimes N}$; and
- A typical sequence can be coded with $H(p)$ bits of classical information per letter which means that a typical sequence $\left|\varphi_{i_1}\right\rangle \otimes \left|\varphi_{i_2}\right\rangle \otimes ... \otimes \left|\varphi_{i_N}\right\rangle$ can be coded with $H(q) = S(\rho)$ qubits of quantum information per state $\left|\varphi_{i_j}\right\rangle$.

- Schumacher's Theorem thus provides a quantum interpretation of the von Neumann entropy $S(\rho)$ that parallels the interpretation of the Shannon entropy $H(p)$ provided by Shannon's Noiseless Coding Theorem.

A new interpretation of logical entropy: I

- In Shannon's statistical rendition of the entropy formula, the use of "typical sequences" is a way of applying the law of large numbers in the form:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} x_j = \sum_{i=1}^{n} p_i x_i.$$

- Shannon's Noiseless Coding Theorem supplies a statistical average rendition of the probabilistic definition:

$$H(p) = \sum_i p_i \log_2\left(\frac{1}{p_i}\right) \text{ where } x_i = \log_2\left(\frac{1}{p_i}\right).$$

- Since logical entropy $h(p) = \sum_i p_i (1 - p_i)$ has a similar probabilistic definition, it also can be rendered as a long run statistical average.

A new interpretation of logical entropy: II

- At each step j in repeated independent sampling $u_1u_2...u_N$ of the probability distribution $p = \{p_1, ..., p_n\}$, the probability that the j^{th} result u_j was not u_j is $1 - \Pr(u_i)$ so the average probability of the result being different than it was in the sequence is:

$$\frac{1}{N}\sum_{j=1}^{N}(1 - \Pr(u_i)).$$

- In the long run, the typical sequences will dominate where the i^{th} outcome is sampled p_iN times so that we have:

$$\sum_{j=1}^{N}(1 - \Pr(u_j))/N \approx \sum_{i=1}^{n} p_iN(1 - p_i)/N = h(p).$$

A new interpretation of logical entropy: III

- The logical entropy $h(p) = \sum_i p_i (1 - p_i)$ is usually interpreted as the *two-draw probability of drawing distinct outcomes* from the distribution $p = \{p_1, ..., p_n\}$.
- Now we have a different interpretation of logical entropy as *the average probability of being different*.

Introduction to Irreducible Representations

The example of $H^{\otimes n}$ as giving irreps for $\mathbb{Z}_2^n$

David Ellerman

UCR

May 2012

Symmetry groups and their irreducible representations

- Group representation theory has important applications in quantum mechanics.
- Given the symmetry group of a certain model, what are the distinct eigen-alternatives compatible with the symmetries?
- As usual, the distinct eigen-alternatives are determined by the 'discrete' or nondegenerate join of partitions that distinguish the alternatives, and the partitions are given by a complete set of commuting operators (CSCO).
- In group representation theory, the distinct eigen-alternatives are the *irreducible representations* (irreps) and their carriers, the *irreducible subspaces*.
- In particle physics, "an elementary particle 'is' an irreducible unitary representation of the group, G of physics,..." [Sternberg, *Group Theory and Physics*, 1994]

Columns of Hadamard Tensor Powers as Irreps

The 2^n columns (or rows) of the $2^n \times 2^n$ matrix $H^{\otimes n}$, which are the $1, -1$-recoded 2^n parity functions (normalized) for the subsets of an n-element set $\{a, b, ..., c\}$, are also the irreps for $\mathbb{Z}_2^n \cong \wp(\{a, b, ..., c\})$ when viewed as additive (Abelian) group:

$$(x_1, ..., x_n) + (y_1, ..., y_n) = (x_1 +_2 y_1, ..., x_n +_2 y_n).$$

In terms of subsets in $\wp(\{a, b, ..., c\})$, the group operation is the symmetric difference operation on subsets.
Hence we can use $H^{\otimes n}$ as a tool to introduce irreps and to see irreps as the eigen-alternatives determined by certain CSCOs.

Group ops as linear operators on group space

- The vector space of functions $V(\mathbb{Z}_2^n) = \{\mathbb{Z}_2^n \to \mathbb{C}\}$ is variously called the *group algebra* or *group space* of $\mathbb{Z}_2^n$ with *standard basis* $|x\rangle$ for $x = (x_1, ..., x_n) \in \mathbb{Z}_2^n$.
- We write the group op multiplicatively so as not to confuse it with formal sums in the group space.
- Each group element $s \in \mathbb{Z}_2^n$ defines a linear operator $T_s : V(\mathbb{Z}_2^n) \to V(\mathbb{Z}_2^n)$ which is defined applying the group operation s to the standard basis vectors: $T_s(|x\rangle) = |s \cdot x\rangle$ (where $s \cdot x$ is the group operation in $\mathbb{Z}_2^n$), and then extended linearly to the whole space.
- The linear operators represented as matrices are just permutation matrices applied to the standard basis vectors (since the group operation carries one group element to another and each group element defines a basis vector).

- For $n = 1$, the group is $\mathbb{Z}_2 \cong \wp(1)$ which, taken multiplicatively, is permutation group S_2.
- $V(\mathbb{Z}_2) = \{\mathbb{Z}_2 \to \mathbb{C}\} \cong \mathbb{C}^2$ is the two-dim. space over $\mathbb{C}$ where the standard basis vectors are $|0\rangle$ and $|1\rangle$.
- The linear op T_0 is the identity and T_1 just interchanges the basis vectors.

$$T_0 = \begin{array}{c|c|c|} & |0\rangle & |1\rangle \\ \hline |0\rangle & 1 & 0 \\ \hline |1\rangle & 0 & 1 \\ \hline \end{array} \text{ and } T_1 = \begin{array}{c|c|c|} & |0\rangle & |1\rangle \\ \hline |0\rangle & 0 & 1 \\ \hline |1\rangle & 1 & 0 \\ \hline \end{array}$$

- The operators commute so the join of the eigenspace partitions is defined.

- For non-Abelian groups, we would have to consider sums of such operators for conjugacy classes of group elements but we can skip those complications by sticking to the Abelian groups $\mathbb{Z}_2^n$ in this introductory treatment.
- The identity operator always has the whole space as eigenspace with eigenvalue $\lambda = 1$ (indiscrete partition).
- The interchange operator T_1 has one eigenspace $\left\{[1,1]^t\right\}$ with $\lambda = 1$ and another eigenspace $\left\{[1,-1]^t\right\}$ with $\lambda = -1$.
- The join of the two vector space partitions is nondegenerate.
- The simultaneous eigenvectors (normalized with positive top entry) are the columns (or rows) of the Hadamard matrix $H = H^{\otimes 1}$.

Application to n = 1 variable functions

- Consider the vector space of functions $f : \mathbb{R} \rightarrow \mathbb{R}$.
- Define the interchange operation as $f(x) \longmapsto f(-x)$ in addition to the identity $f(x) \longmapsto f(x)$ so that S_2 operates on that space of functions.
- The irreps provide a vector space partition of the space into the:
  1. Even functions $\frac{1}{2}[f(x) + f(-x)]$, and the
  2. Odd functions $\frac{1}{2}[f(x) - f(-x)]$.
- Per usual with a vector space partition, an arbitrary function can be decomposed into a unique sum of an even function and an odd function:

$$f(x) = \frac{1}{2}[f(x) + f(-x)] + \frac{1}{2}[f(x) - f(-x)].$$

- Not coincidentally, the two types of functions are associated with the two types of parity.

Simplest non-trivial example of n = 2: I

- For $n = 2$, the group is $\mathbb{Z}_2^2 \cong \wp(2)$, which taken multiplicatively is the Klein four-group as well as the dihedral group D_2.
- We represent a symmetry group like D_2 as starting with an original configuration such as ${}_{c}^{b}\square_{d}^{a}$ and then to represent each symmetry operation by the resulting configuration. For instance, the rotation C_2 around 180° gives the configuration ${}_{a}^{d}\square_{b}^{c}$. The dihedral group D_2 also has the symmetry operations: C_2^h = flipping around the horizontal axis and C_2^v = flipping around the vertical axis. If we think of the top side as the light square, then flipping shows the 'dark underside' so C_2^v gives ${}_{d}^{a}\blacksquare_{c}^{b}$ and C_2^h gives ${}_{b}^{c}\blacksquare_{a}^{d}$.
- The group multiplication table is as follows (column op. applied first):

D_2	I	C_2	C_2^h	C_2^v
I	${}^{b}_{c}\square^{a}_{d}$	${}^{d}_{a}\square^{c}_{b}$	${}^{c}_{b}\blacksquare^{d}_{a}$	${}^{a}_{d}\blacksquare^{b}_{c}$
C_2	${}^{d}_{a}\square^{c}_{b}$	${}^{b}_{c}\square^{a}_{d}$	${}^{a}_{d}\blacksquare^{b}_{c}$	${}^{c}_{b}\blacksquare^{d}_{a}$
C_2^h	${}^{c}_{b}\blacksquare^{d}_{a}$	${}^{a}_{d}\blacksquare^{b}_{c}$	${}^{b}_{c}\square^{a}_{d}$	${}^{d}_{a}\square^{c}_{b}$
C_2^v	${}^{a}_{d}\blacksquare^{b}_{c}$	${}^{c}_{b}\blacksquare^{d}_{a}$	${}^{d}_{a}\square^{c}_{b}$	${}^{b}_{c}\square^{a}_{d}$

- Each group operation defines a linear operator on $V\left(\mathbb{Z}_2^2\right) \cong \mathbb{C}^2 \otimes \mathbb{C}^2$ which is represented by a permutation matrix (op. applied to column symbol gives row symbol). The non-identity ones are:

	T_{C_2}	${}^{b}_{c}\square^{a}_{d}$	${}^{d}_{a}\square^{c}_{b}$	${}^{c}_{b}\blacksquare^{d}_{a}$	${}^{a}_{d}\blacksquare^{b}_{c}$
	${}^{b}_{c}\square^{a}_{d}$	0	1	0	0
1	${}^{d}_{a}\square^{c}_{b}$	1	0	0	0
	${}^{c}_{b}\blacksquare^{d}_{a}$	0	0	0	1
	${}^{a}_{d}\blacksquare^{b}_{c}$	0	0	1	0

2

$T_{C_2^h}$	${}^b_c\square^a_d$	${}^d_a\square^c_b$	${}^c_b\blacksquare^d_a$	${}^a_d\blacksquare^b_c$
${}^b_c\square^a_d$	0	0	1	0
${}^d_a\square^c_b$	0	0	0	1
${}^c_b\blacksquare^d_a$	1	0	0	0
${}^a_d\blacksquare^b_c$	0	1	0	0

3

$T_{C_2^v}$	${}^b_c\square^a_d$	${}^d_a\square^c_b$	${}^c_b\blacksquare^d_a$	${}^a_d\blacksquare^b_c$
${}^b_c\square^a_d$	0	0	0	1
${}^d_a\square^c_b$	0	0	1	0
${}^c_b\blacksquare^d_a$	0	1	0	0
${}^a_d\blacksquare^b_c$	1	0	0	0

- These 3 permutation matrices along with the identity commute so we can take the join of their eigenspace partitions.

Example n = 2; Eigenspace partitions: I

- As always, T_I has the indiscrete eigenspace partition with $\lambda = 1$.

1. T_{C_2}: $\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \right\} \lambda = -1; \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \right\} \lambda = 1$
2. $T_{C_2^h}$: $\left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \right\} \lambda = -1; \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \right\} \lambda = 1$
3. $T_{C_2^v}$: $\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \right\} \lambda = -1; \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \right\} \lambda = 1$

Example n = 2; Eigenspace partitions: II

- Join of eigenspace partitions is nondegenerate and the simultaneous eigenvectors are the *irreducible representations* or *irreps* of the group:

$$\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}, \begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}$$

- Normalized (with positive top entry), they are the columns of $H^{\otimes 2}$.

$$H^{\otimes 2} = \frac{1}{2}\begin{bmatrix}1 & 1 & 1 & 1\\1 & -1 & 1 & -1\\1 & 1 & -1 & -1\\1 & -1 & -1 & 1\end{bmatrix}$$

Irreps as maximally distinct eigen-alternatives: I

- The standard basis vectors in $\{\mathbb{Z}_2^2 \rightarrow \mathbb{C}\}$ can be visually represented by the four configurations resulting from the four operations of D_2 which is isomorphic to $\mathbb{Z}_2^2$.
- Then the irreps as the simultaneous eigenvectors of the four commuting operators defined by the group operations are given in the following table along with the n-tuples of eigenvalues that characterize each irrep (where we have also simplified the configuration symbols).

$\chi^{(i)}$	$\left\| \lambda_0^{(i)}, \lambda_1^{(i)}, \lambda_2^{(i)}, \lambda_3^{(i)} \right\rangle$
$\chi^{(0)} = \square^a + \square^c + \blacksquare^d + \blacksquare^b$	$\|1,1,1,1\rangle$
$\chi^{(1)} = \square^a + \square^c - \blacksquare^d - \blacksquare^b$	$\|1,1,-1,-1\rangle$
$\chi^{(2)} = \square^a - \square^c + \blacksquare^d - \blacksquare^b$	$\|1,-1,1,-1\rangle$
$\chi^{(3)} = \square^a - \square^c - \blacksquare^d + \blacksquare^b$	$\|1,-1,-1,1\rangle$

Irreps as maximally distinct eigen-alternatives: II

- Note firstly that each subspace generated by the irreps is invariant under all the operations (which we already know since they are *simultaneous* eigenvectors).
- For instance, apply C_2, the 180° rotation, to $\chi^{(2)} = \square^a - \square^c + \blacksquare^d - \blacksquare^b$ to get:

$$C_2\left(\chi^{(2)}\right) = \square^c - \square^a + \blacksquare^b - \blacksquare^d = -1\chi^{(2)} \text{ where } \lambda_1^{(2)} = -1.$$

- But the whole space or any subspace generated by subsets of the irreps is also invariant.
- The distinctive feature of the irreps is they are the *minimal* invariant subspaces which means they are *maximally distinguished* invariant subspaces. They are carved out by the (nondegenerate) join of the eigenspace partitions as indicated by the ordered n-tuples of their eigenvalues.

Irreps as maximally distinct eigen-alternatives: III

- Just as an element of a set may be maximally distinguished from other elements by its values in a complete set of attributes, e.g., height, weight, age,..., so each irrep is characterized by the ordered n-tuple of eigenvalues for the complete set of commuting operators (CSCO) for the group operations.
- Thus $\chi^{(2)}$ is characterized by its eigenvalue labels $|1, -1, 1, -1\rangle$ which are the values of the irrep "eigenstate" under the "observable" operators T_I, T_{C_2}, $T_{C_2^h}$, and $T_{C_2^v}$.
- With only two eigenvalues, a yes-or-no game of "20 questions" characterizes $\chi^{(2)}$:
  1. Question 1: For observable T_I, is $\lambda = 1$? Yes.
  2. Question 2: For observable T_{C_2}, is $\lambda = 1$? No.
  3. Question 3: For observable $T_{C_2^h}$, is $\lambda = 1$? Yes.
  4. Question 4: For observable $T_{C_2^v}$, is $\lambda = 1$? No.

Vector space of 2-variable functions $V = \{f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}\}$. Define the group operations: I: $f(x,y) \longmapsto f(x,y)$; C_2: $f(x,y) \longmapsto f(x,-y)$; C_2^h: $f(x,y) \longmapsto f(-x,y)$; C_2^v: $f(x,y) \longmapsto f(-x,-y)$. This is a *group representation* $D \to Lin(V,V)$ if it satisfies group mult. table.

D_2	I	C_2	C_2^h	C_2^v
I	$f(x,y)$	$f(x,-y)$	$f(-x,y)$	$f(-x,-y)$
C_2	$f(x,-y)$	$f(x,y)$	$f(-x,-y)$	$f(-x,y)$
C_2^h	$f(-x,y)$	$f(-x,-y)$	$f(x,y)$	$f(x,-y)$
C_2^v	$f(-x,-y)$	$f(-x,y)$	$f(x,-y)$	$f(x,y)$

Then the irreps define the four types of functions:

1. $\frac{1}{4}[f(x,y) + f(x,-y) + f(-x,y) + f(-x,-y)]$; x, y even;
2. $\frac{1}{4}[f(x,y) - f(x,-y) + f(-x,y) - f(-x,-y)]$; x even, y odd;
3. $\frac{1}{4}[f(x,y) + f(x,-y) - f(-x,y) - f(-x,-y)]$; x odd, y even;
4. $\frac{1}{4}[f(x,y) - f(x,-y) - f(-x,y) + f(-x,-y)]$; x odd, y odd.

Application to n = 2 variable functions: II

- Per usual with partitions, any function $f(x,y)$ is uniquely the sum of functions of the four types.
- Not coincidentally, the four types are associated with the four parity basis functions χ_s for $n = 2$.
- If $s = (0,0), (0,1), (1,0)$, or $(1,1)$, then the parity function χ_s gives the sign pattern, e.g., for $s = (0,1)$, $\chi_{(0,1)}$ gives the pattern $(+,-,+,-)$ for the second type of symmetrized function $\frac{1}{4}[f(x,y) - f(x,-y) + f(-x,y) - f(-x,-y)]$.
- The general case for any n gives the 2^n distinct eigen-alternatives types of parity, e.g., to classify the parity types of n-variable functions $f(x_1, ..., x_n)$.
- Those 2^n types of parity are given by the sign patterns in the columns of $H^{\otimes n}$.