Introduction to Quantum Logical Information Theory: Talk

David Ellerman¹,a

¹University of California Riverside

Abstract. Logical information theory is the quantitative version of the logic of partitions just as logical probability theory is the quantitative version of the dual Boolean logic of subsets. The resulting notion of information is about distinctions, differences, and distinguishability, and is formalized using the distinctions ("dits") of a partition (a pair of points distinguished by the partition). All the definitions of simple, joint, conditional, and mutual entropy of Shannon information theory are derived by a uniform transformation from the corresponding definitions at the logical level. The purpose of this talk is to outline the direct generalization to quantum logical information theory that similarly focuses on the pairs of eigenstates distinguished by an observable, i.e., "qudits" of an observable. The fundamental theorem for quantum logical entropy and measurement establishes a direct quantitative connection between the increase in quantum logical entropy due to a projective measurement and the eigenstates (cohered together in the pure superposition state being measured) that are distinguished by the measurement (decohered in the post-measurement mixed state). Both the classical and quantum versions of logical entropy have simple interpretations as “two-draw” probabilities for distinctions. The conclusion is that quantum logical entropy is the simple and natural notion of information for quantum information theory focusing on the distinguishing of quantum states.

1 Introduction

Subsets and partitions are category-theoretic dual concepts. The elements of subsets are dual to the distinctions (dits) of a partition, i.e., an ordered pair of elements in different blocks. Thus the usual Boolean logic of subsets (often taken as “propositional logic”) has a dual logic of partitions [2] where elements of subsets and distinctions of partition have corresponding roles. Gian-Carlo Rota said

Figure 1. Partition $\pi = \{B_1, ..., B_6\}$ on set $U = \{u_1, ..., u_n\}$

Boolean logic of subsets (often taken as “propositional logic”) has a dual logic of partitions [2] where elements of subsets and distinctions of partition have corresponding roles. Gian-Carlo Rota said

⁎e-mail: david@ellerman.org
Figure 2. Dual Logics: Boolean subset logic and partition logic.

“Probability is a measure on the Boolean algebra of events” that gives quantitatively the “intuitive idea of the size of a set”, so we may ask by “analogy” for some measure to capture a property for a partition like “what size is to a set.” Rota goes on to ask:

How shall we be led to such a property? We have already an inkling of what it should be: it should be a measure of information provided by a random variable. Is there a candidate for the measure of the amount of information? [7, p. 67]

Since elements are to subsets as distinctions are to partitions, the “size” of a partition may be taken as the number of distinctions.

2 Logical Information Theory

The new logical foundations for information theory starts with sets, not probabilities, as suggested by Andrei Kolmogorov.

Information theory must precede probability theory, and not be based on it. By the very essence of this discipline, the foundations of information theory have a finite combinatorial character. [4, p. 39]

The notion of information-as-distinctions thus starts with the set of distinctions, the information set, of a partition \( \pi = \{B, B', \ldots\} \) on a finite set \( U \) where that set of distinctions (dits) is:

\[
dit(\pi) = \{ (u, u') : \exists B, B' \in \pi, B \neq B', u \in B, u' \in B' \}. \tag{1}
\]

The ditset of a partition is the complement in \( U \times U \) of the equivalence relation associated with the partition \( \pi \). Given any probability measure \( p : U \to [0, 1] \) on \( U = \{u_1, \ldots, u_n\} \) which defines \( p_i = p(u_i) \) for \( i = 1, \ldots, n \), the product measure \( p \times p : U \times U \to [0, 1] \) has for any \( S \subseteq U \times U \) the value of:

\[
p \times p(S) = \sum_{(u_i, u_j) \in S} p(u_i) p(u_j) = \sum_{(u_i, u_j) \in S} p_i p_j. \tag{2}
\]

The logical entropy of \( \pi \) is the product measure of its ditset:

\[
h(\pi) = p \times p(dit(\pi)) = \sum_{(u_i, u_j) \in \text{dit}(\pi)} p_i p_j = 1 - \sum_{B \in \pi} p(B)^2. \tag{3}
\]
Given partitions $\pi = \{B_1, ..., B_l\}, \sigma = \{C_1, ..., C_j\}$ on $U$, the information set or ditset for their join is:

$$
dit(\pi \cup \sigma) = dit(\pi) \cup dit(\sigma) \subseteq U \times U. \tag{4}
$$

Given probabilities $p = \{p_1, ..., p_n\}$, the joint logical entropy is:

$$
h(\pi, \sigma) = h(\pi \cup \sigma) = p \times p(dit(\pi) \cup dit(\sigma)) = 1 - \sum_{i,j} p(B_i \cap C_j)^2. \tag{5}
$$

The info set for the conditional logical entropy $h(\pi|\sigma)$ is the difference of dit sets, and thus: $h(\pi|\sigma) = p \times p(dit(\pi) - dit(\sigma))$.

The info set for the logical mutual information $m(\pi, \sigma)$ is the intersection of dit sets, so: $m(\pi, \sigma) = p \times p(dit(\pi) \cap dit(\sigma))$. The Information algebra $I(\pi, \sigma) = \text{Boolean subalgebra of } \varphi(U \times U)$ generated by dit sets and their complements.

### 3 Deriving the Shannon entropies from the logical entropies

The simple and compound definitions for Shannon entropy are often presented as satisfying a Venn diagram, but they are not a measure (in the sense of measure theory). However, all the Shannon entropies can be derived from the definitions of logical entropy (which is a measure) by a uniform transformation that preserves the Venn diagram relationships. That “dit-bit transformation” starts with logical entropy measured in dits and requantifies in terms of bits to obtain the corresponding Shannon entropy.

In the canonical case of $n$ equiprobable elements, $p_i = \frac{1}{n}$, the logical entropy of $p = \{\frac{1}{n}, ..., \frac{1}{n}\}$ is:

$$
\frac{|U \times U - \Delta|}{|U \times U|} = \frac{n^2 - n}{n^2} = 1 - \frac{1}{n} = 1 - p_i. \tag{6}
$$
In the general case \( p = (p_1, \ldots, p_n) \), the logical entropy is the average of this dit-count \( 1 - p_i \):

\[
h(p) = \sum_{i=1}^{n} p_i (1 - p_i).
\] (7)

In the canonical case of \( 2^n \) equiprobable elements so \( p_i = \frac{1}{2^n} \), the minimum number of binary partitions ("yes-or-no questions") or "bits" it takes to uniquely determine or encode each distinct element or block is \( n \), so the Shannon-Hartley entropy is:

\[
n = \log_2 (2^n) = \log_2 \left( \frac{1}{\frac{1}{2^n}} \right) = \log_2 \left( \frac{1}{p_i} \right).
\] (8)

In the general case of \( p = (p_1, \ldots, p_{2^n}) \), the Shannon entropy is the average of this bit-count \( \log_2 \left( \frac{1}{p_i} \right) \):

\[
H(p) = \sum_{i=1}^{2^n} p_i \log_2 \left( \frac{1}{p_i} \right).
\] (9)

Hence the Dit-Bit Transform is: express any logical entropy concept (joint, conditional, or mutual) as average of dit-counts \( 1 - p_i \), and then substitute the bit-count \( \log_2 \left( \frac{1}{p_i} \right) \) to obtain the corresponding formula as defined by Shannon. For the corresponding definitions for random variables and their probability distributions, consider a random variable \((x, y)\) taking values on the product \(X \times Y\) of finite sets with the joint probability distribution \( \{p_{xy}\} \), and thus with the marginal distributions \( \{p_x\} \) and \( \{p_y\} \) where \( p_x = \sum_{y \in Y} p_{xy} \) and \( p_y = \sum_{x \in X} p_{xy} \). The entropies can be considered as functions of the random variables or of their probability distributions, e.g., \( h\left( \{p_{xy}\} \right) = h(x, y) \), \( h\left( \{p_x\} \right) = h(x) \), and \( h\left( \{p_y\} \right) = h(y) \). The dit-bit transform preserves the same Venn diagram formulas for the Shannon entropies in spite of them not being a measure (in the sense of measure theory).

| Table 3 | The Dit-Bit Transform: 1- p_i → log(1/p_i) |
|-----------------------------------------------|
| **Entropy** | \( h(p) = \sum p_i (1 - p_i) \) |
| **Joint Entropy** | \( h(x, y) = \sum_{x, y} p_{xy} (1 - p_{xy}) \) |
| **Conditional Entropy** | \( h(x|y) = \sum_{x, y} p_{xy} \log(1/p_{xy}) \) |
| **Mutual Information** | \( I(x;y) = \sum_{x, y} p_{xy} (\log(1/p_{xy}) - \log(1/p_x)) \) |

**Figure 5.** Venn-diagram-preserving dit-bit transform.

**Figure 6.** Venn diagram ‘mnemonic’ for Shannon entropies.
4 Logical entropy via density matrices

All this will carry over to quantum logical entropy using density matrices. ‘Classically,’ the density matrix representing the event $S$ is the $n \times n$ symmetric real matrix:

$$\rho(S) = \rho(S) = |S\rangle \langle S| = \begin{cases} \frac{1}{|S|} \sqrt{P_{ij} P_{kl}} & \text{for } u_j, u_k \in S \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Then $\rho(S)^2 = |S\rangle \langle S| \langle S| = \rho(S)$ so borrowing language from QM, $|S\rangle$ is said to be a pure state or event. Given any partition $\pi = \{B_1, ..., B_s\}$ on $U$, its density matrix is the average of the block density matrices:

$$\rho(\pi) = \sum_i \rho(B_i) \rho(B_i). \quad (11)$$

Then $\rho(\pi)$ represents the mixed state, experiment, or lottery where the event $B_i$ occurs with probability $p(B_i)$.

**Example 1** For the throw of a fair die, $U = \{u_1, u_3, u_5, u_2, u_4, u_6\}$ (where $u_j$ represents the number $j$ coming up), the density matrix $\rho(0_U)$ is the “pure state” $6 \times 6$ matrix with each entry being $\frac{1}{6}$ (note the odd states listed first)

$$\rho(0_U) = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix} u_1$$

The nonzero off-diagonal entries represent indistinctions or indits of partition $0_U$, or in quantum terms as “coherences” where all 6 “eigenstates” cohere together in a pure “superposition” state. All pure states have logical entropy of zero, i.e., $h(0_U) = 0$ (i.e., no dits).

**Example 2 (continued)** Now classify or “measure” the elements by parity (odd or even) partition (observable) $\pi = \{B_{\text{odd}}, B_{\text{even}}\} = \{\{u_1, u_3, u_5\}, \{u_2, u_4, u_6\}\}$. Mathematically, this is done by the Lüders mixture operation where $P_{\text{odd}}$ and $P_{\text{even}}$ are the projections to the odd or even components:

$$P_{\text{odd}} \rho(0_U) P_{\text{odd}} + P_{\text{even}} \rho(0_U) P_{\text{even}} = \sum_{i=1}^m p(B_i) \rho(B_i) = \rho(\pi). \quad (13)$$

$$\rho(0_U) = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix} \Rightarrow \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \end{bmatrix} = \rho(\pi) \quad (14)$$

**Theorem 3 (Basic)** The increase in logical entropy due to a Lüders mixture operation is the sum of amplitudes squared of the non-zero off-diagonal entries of the beginning density matrix that are zeroed in the final density matrix.
Proof. Since for any density matrix \( \rho \), \( tr[\rho^2] = \sum_{i,j} |\rho_{ij}|^2 \) [3, p. 77], \( h(\rho(\pi)) - h(\rho(0_U)) = (1 - tr[\rho(\pi)^2]) - (1 - tr[\rho(0_U)^2]) \) = \( tr[\rho(0_U)^2] - tr[\rho(\pi)^2] = \sum_{i,j} |\rho_{ij}(0_U)|^2 - \sum_{i,j} |\rho_{ij}(\pi)|^2 \). If \((u_i, u_{i'}) \in \text{dit}(\pi)\), then and only then are the off-diagonal terms corresponding to \(u_i\) and \(u_{i'}\) zeroed by the Lüders operation. \[ \Box \]

Example 4 (continued) In comparison with the matrix \( \rho(0_U) \) of all entries \( \frac{1}{6} \), the entries that got zeroed in \( \rho(0_U) \sim \rho(\pi) \) correspond to the distinctions created in the transition \( 0_U \sim \pi = \{\{u_1, u_3, u_5\}, \{u_2, u_4, u_6\}\} \). Increase in logical entropy \( = h(\pi) - h(0_U) = 2 \times 9 \times (\frac{1}{6})^2 = \frac{18}{36} = \frac{1}{2} \). Usual calculations: \( h(\pi) = 1 - 2 \times (\frac{1}{2})^2 = \frac{1}{2} \text{ and } h(0_U) = 1 - 1^2 = 0 \).

Since a projective measurement’s effect on a density matrix in QM is the Lüders mixture operation, that means that the effects of the measurement is the above-described “making distinctions” by decohering or zeroing certain coherence terms in the density matrix, and the sum of the absolute squares of the coherences that were decohered is the change in the logical entropy.

5 Generalization to quantum logical information theory

The logical notion of information-as-distinctions generalizes to quantum information theory.

[Information] is the notion of distinguishability abstracted away from what we are distinguishing, or from the carrier of information. ...And we ought to develop a theory of information which generalizes the theory of distinguishability to include these quantum properties... [1, p. 155]

A quantum bit, a qubit, or rather a qudit, needs to be relativized to an observable–just as a dit is a dit of a partition.

A qubit of an observable \( F \) is a pair \((u_k, u_{k'})\) in the eigenbasis definitely distinguishable by \( F \), i.e., \( \phi(u_k) \neq \phi(u_{k'}) \), distinct eigenvalues. Since the quantum version of logical entropy is a straightforward generalization of sets to vector spaces, we give the generalization in a table form–first before the introduction of probabilities.

<table>
<thead>
<tr>
<th>Classical Logical Info. Theory</th>
<th>Quantum Logical Info. Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 4a (w/o probes):</strong></td>
<td><strong>Orthogonal basis of Hilbert space ( \mathbb{H} )</strong></td>
</tr>
<tr>
<td>Attribute/Observable</td>
<td>( U = {u_1, u_2, u_3, u_4} )</td>
</tr>
<tr>
<td>Real-valued random variables</td>
<td>Composing self-adjoint operators ( F ), ( G ) ( {u}, \text{O.N. basis of sum of eigenvalues} )</td>
</tr>
<tr>
<td>Values</td>
<td><strong>Eigenvectors</strong> of ( F ) ( {u} ) <strong>Eigenvectors</strong> of ( G ) ( {u} )</td>
</tr>
<tr>
<td>Inverse-image</td>
<td><strong>Image values</strong> of ( F ) ( {u} ) <strong>Image values</strong> of ( G ) ( {u} )</td>
</tr>
<tr>
<td>Distinctions</td>
<td><strong>Inverses</strong> of ( F ) ( {u} ) <strong>Inverses</strong> of ( G ) ( {u} )</td>
</tr>
<tr>
<td>Information set/space</td>
<td><strong>Quotient</strong> of ( F ) ( {u} ) <strong>Quotient</strong> of ( G ) ( {u} )</td>
</tr>
<tr>
<td>Deterministic Mutual</td>
<td><strong>Quotient</strong> of ( F ) ( {u} ) <strong>Quotient</strong> of ( G ) ( {u} )</td>
</tr>
</tbody>
</table>

Figure 7. The generalization of classical to quantum logical entropies before probabilities.

The fundamental theorem connecting measurement to logical entropy carries over to the quantum case. The nonzero off-diagonal terms in density matrix \( \rho(\psi) \) are called “coherences”–like indistinctions in classical case. Measurement creates distinctions, i.e., turn coherences into ‘decoherences’–classically, turn indistinctions into distinctions.
Figure 8. The generalization of classical to quantum logical entropies with probabilities.

**Basic Theorem**: Measure of distinctions created in measuring pure state $\psi$ by $F = \text{sum of absolute squares of off-diagonal terms zeroed (i.e., coherences that were decohered)}$ in measurement = logical entropy increase, e.g., $h (F : \psi) = h (\rho^F (\psi)) = h (\rho (\psi)) = \text{probability that two independent measurements of } \psi \text{ will yield a qudit of } F$.

This treatment can be generalized to the case of two non-commuting observables and a given state. There is also the case where we are only given density operators $\rho$ and $\tau$. Then the logical entropy of a state $\rho$ is $h (\rho) = 1 - \text{tr} [\rho^2]$ and the quantum logical cross-entropy of $\rho$ and $\tau$ is: $h (\rho || \tau) = 1 - \text{tr} [\rho \tau]$. Then the classical notion of Hamming distance can be generalized to the quantum logical Hamming distance between the two states:

$$d (\rho, \tau) = 2h (\rho || \tau) - h (\rho) - h (\tau).$$

This treatment avoids the problem suggested by Nielsen and Chuang.

"Unfortunately, the Hamming distance between two objects is simply a matter of labeling, and a priori there aren’t any labels in the Hilbert space arena of quantum mechanics!" [6, p. 399]

Then the main result is that the quantum logical Hamming distance is same as the usual Hilbert-Schmidt distance [8] (some authors include a factor of 1/2) whose square root is the trace-distance.

Quantum logical Hamming distance $d (\rho, \tau) = \text{tr} [(\rho - \tau)^2]$ Hilbert-Schmidt distance.

**References**


