Inequalities Witnessing Quantum Incompatibility in The Triangle Scenario

Thomas C. Fraser*

Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada

University of Waterloo, Waterloo, Ontario, Canada

(Dated: August 18, 2016)

This document is my current working draft of a paper to do with causal inference, inflation, incompatibility inequalities, hypergraph transversals and quantum correlations.

* tcfraser@tcfraser.com

I. INTRODUCTION

II. DEFINITIONS & NOTATION

Definition 1. Borrowing the notation from [1], each random variable v has a set of all possible outcomes called the **outcome space** or **valuation space** and is denoted O_v. When referencing a *specific* element of O_v or *valuation* of v, the notation $o[v]$ is used. This notation generalizes to set of random variables $V = \{v_1, \ldots, v_{|V|}\}$; a specific outcome of $o[V] \in O_V$ is used to reference a particular tuple or vector of outcomes,

$$o[V] \equiv \left(o[v_1], o[v_2], \ldots, o\left[v_{|V|}\right]\right) = (o[v])_{v \in V} \tag{1}$$

It is important to note that the ordering of events in eq. (1) is irrelevant. Similarly, the outcome space O_V for a set of random variables V is the **valuation product** of the individual outcome spaces,

$$O_V \equiv O_{v_1} \times \cdots \times O_{v_{|V|}}$$

Where '$\times$' denotes is defined such that,

$$o[A] \times o[B] \equiv \left(o[a_1], \ldots, o\left[a_{|A|}\right], o[b_1], \ldots, o\left[b_{|B|}\right]\right)$$

When referencing the probability that and set of variables V has outcome $o[V]$ (say $v_1 = 0, v_2 = 3$), we intend for the following class of notations to be used interchangeably.

$$P_V(o[V]) = P(o[V]) = P(o[v_1]\, o[v_2]) = P(v_1 = 0, v_2 = 3) = P(v_2 = 3, v_1 = 0) = P_{v_1, v_2}(03)$$

Definition 2. An outcome $o[V]$ is said to be **extendable** to an outcome $o[W]$ (where $V \subseteq W$) if there exists an outcome $o[W \setminus V]$ such that:

$$o[W] = o[W \setminus V] \times o[V]$$

The idea being that a *less specific* outcome $o[V]$ can be made *more specific* by assigning outcomes to the remaining random variables in $W \setminus V$. If such an outcome $o[W \setminus V]$ exists, it is unique.

Definition 3. The set of all extendable outcomes of $o[V]$ in O_W is called the **extendable set** and can be written as,

$$o[V] \times O_{W \setminus V} \equiv \{o[W] \in O_W \mid \exists o[W \setminus V] : o[W \setminus V] \times o[V] = o[W]\}$$

The extendable set of $o[V]$ in W is the set of all outcomes of O_W that *agree* with $o[V]$ about valuations for variables in V. In this language, two outcomes $o[V]$ and $o[W]$ are **compatible** if they are both extendable to some outcome $o[V \cup W]$. Equivalently, $o[V]$ and $o[W]$ agree on their valuations of $V \cap W$.

Example 4. Consider two sets of random variables $V = \{a, b\}$ and $W = \{a, b, c\}$. Clearly $V \subseteq W$; a prerequisite for extendability. Also take all individual outcome spaces to be finite and of order 3: $O_a = O_b = O_c = \{1, 2, 3\}$. Then $o[V] = o[\{a, b\}] = (a = 1, b = 2)$ is extendable to the outcome $o[W] = (a = 1, b = 2, c = 1)$, and the extendable set of $o[V]$ in O_W is,

$$o[\{a, b\}] \times O_c = \{(a = 1, b = 2, c = 1), (a = 1, b = 2, c = 2), (a = 1, b = 2, c = 3)\}$$

Definition 5. A **graph** is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *edges* respectively where the nodes can represent any object and the edges are pairs of nodes. For convenience of notation, one defines an index set over the nodes denoted $\mathcal{I}_\mathcal{N}$.

$$\mathcal{N} = \{n_i \mid i \in \mathcal{I}_\mathcal{N}\} \quad \mathcal{E} = \{\{n_j, n_k\} \mid j, k \in \mathcal{I}_\mathcal{N}\}$$

Definition 6. A **directed graph** $\mathcal{G}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *edges* respectively where the nodes can represent any object and the edges are *ordered* pairs of nodes. For convenience of notation, one defines an index set over the nodes denoted $\mathcal{I}_\mathcal{N}$.

$$\mathcal{N} = \{n_i \mid i \in \mathcal{I}_\mathcal{N}\} \quad \mathcal{E} = \{n_j \to n_k \mid j, k \in \mathcal{I}_\mathcal{N}\}$$

Definition 7. The following definitions are common language in directed graph theory. Let $n, m \in \mathcal{N}$ be example nodes of the graph $\mathcal{G}$.

- The **parents of a node**: $\mathrm{Pa}_{\mathcal{G}}(n) \equiv \{m \mid m \to n\}$
- The **children of a node**: $\mathrm{Ch}_{\mathcal{G}}(n) \equiv \{m \mid n \to m\}$
- The **ancestry of a node**: $\mathrm{An}_{\mathcal{G}}(n) \equiv \bigcup_{i \in \mathbb{W}} \mathrm{Pa}^{i}_{\mathcal{G}}(n)$ where $\mathrm{Pa}^{i}_{\mathcal{G}}(n) \equiv \mathrm{Pa}_{\mathcal{G}}\big(\mathrm{Pa}^{i-1}_{\mathcal{G}}(n)\big)$ and $\mathrm{Pa}^{0}_{\mathcal{G}}(n) = n$

All of these terms can be generalized to sets of nodes $N \subseteq \mathcal{N}$ through union over the elements,

- The **parents of a node set**: $\mathrm{Pa}_{\mathcal{G}}(N) \equiv \bigcup_{n \in N} \mathrm{Pa}_{\mathcal{G}}(n)$
- The **children of a node set**: $\mathrm{Ch}_{\mathcal{G}}(N) \equiv \bigcup_{n \in N} \mathrm{Ch}_{\mathcal{G}}(n)$
- The **ancestry of a node set**: $\mathrm{An}_{\mathcal{G}}(N) \equiv \bigcup_{n \in N} \mathrm{An}_{\mathcal{G}}(n)$

Moreover, an **induced subgraph** of $\mathcal{G}$ due to a set of nodes $N \subseteq \mathcal{N}$ is the graph composed of N and all edges $e \in \mathcal{E}$ of the original graph that are contained in N.

$$\mathrm{Sub}_{\mathcal{G}}(N) \equiv (N, \{e_i \mid i \in \mathcal{I}_{\mathcal{E}}, e_i \subseteq N\})$$

An **ancestral subgraph** of $\mathcal{G}$ due to $N \subseteq \mathcal{N}$ is the induced subgraph due to the ancestry of N.

$$\mathrm{AnSub}_{\mathcal{G}}(N) \equiv \mathrm{Sub}_{\mathcal{G}}\big(\mathrm{An}_{\mathcal{G}}(N)\big)$$

Definition 8. A **directed acyclic graph** or **DAG** $\mathcal{G}$ is an directed graph definition 6 with the additional property that no node n is in its set of **ancestors**.

$$\forall n \in \mathcal{N} : n \notin \bigcup_{i \in \mathbb{N}} \mathrm{Pa}^{i}_{\mathcal{G}}(n)$$

Notice the difference between using the natural numbers $\mathbb{N}$ to distinguish *ancestors* from *ancestry*.

Definition 9. A **hypergraph** denoted $\mathcal{H}$ is an ordered tuple $(\mathcal{N}, \mathcal{E})$ of *nodes* and *edges* respectively where the nodes can represent any object and the edges are *subsets* of nodes. For convenience of notation, one defines an index set over the nodes and edges of a hypergraph $\mathcal{H}$ denoted $\mathcal{I}_{\mathcal{N}}$ and $\mathcal{I}_{\mathcal{E}}$ respectively.

$$\mathcal{H} = (\mathcal{N}, \mathcal{E}) \quad \mathcal{N} = \{n_i \mid i \in \mathcal{I}_{\mathcal{N}}\} \quad \mathcal{E} = \{e_i \mid i \in \mathcal{I}_{\mathcal{E}}, e_i \subseteq \mathcal{N}\}$$

Note that whenever the index for an edge or node is arbitrary, it will be omitted. There is a dual correspondence between edges $e \in \mathcal{E}$ and nodes $n \in \mathcal{N}$ in a Hypergraph. An edge e is viewed as a set of nodes $\{n_i\}$, and a node n can be viewed as the set of edges $\{e_i\}$ that contain it.

Definition 10. A **hypergraph transversal** (or edge hitting set) $\mathcal{T}$ of a hypergraph $\mathcal{H}$ is a set of nodes $\mathcal{T} \subseteq \mathcal{N}$ that have non-empty intersections with every edge in $\mathcal{E}$.

$$\mathcal{T} = \{n_i \in \mathcal{N} \mid i \in \mathcal{I}_{\mathcal{T}}\} \quad \forall e \in \mathcal{E} : \mathcal{T} \cap e \neq \emptyset$$

Definition 11. A **minimal hypergraph transversal** $\mathcal{T}$ is any valid transversal (definition 10) of $\mathcal{H}$ where every node n is *necessary* to retain validity. For each node n in $\mathcal{T}$, $\mathcal{T} \setminus n$ is no longer a transversal.

$$\mathcal{T} = \{n_i \in \mathcal{N} \mid i \in \mathcal{I}_{\mathcal{T}}\} \quad \forall i \in \mathcal{I}_{\mathcal{T}}, \exists e \in \mathcal{E} : (\mathcal{T} \setminus n_i) \cap e = \emptyset$$

Definition 12. A **weighted hypergraph** $\mathcal{H}_{\mathcal{W}}$ is a regular hypergraph satisfying definition 9 equipped with a set of weights $\mathcal{W}$ ascribed to each node such that a weighted hypergraph is written as a triplet $(\mathcal{W}, \mathcal{N}, \mathcal{E})$.

$$\mathcal{W} = \{w_i \mid i \in \mathcal{I}_{\mathcal{N}}, w_i \in \mathbb{R}\}$$

One would say that a particular node n_i carries weight w_i for each $i \in \mathcal{I}_{\mathcal{N}}$.

Definition 13. A **bounded transversal** of a weighted hypergraph $\mathcal{H}_{\mathcal{W}}$ is a transversal $\mathcal{T}$ of the unweighted hypergraph $\mathcal{H}$ and a real number t (denoted $\mathcal{T}_{\leq t}$) such that the sum of the node weights of the transversal is bounded by t.

$$\mathcal{T}_{\leq t} = \{n_i \mid i \in \mathcal{I}_{\mathcal{T}}\} \quad \text{s.t.} \sum_{j \in \mathcal{I}_{\mathcal{T}}} w_j \leq t$$

One can definte analogous **(strictly) upper/lower bounded transversals** by considering modifications of the notation: $\mathcal{T}_{<t}, \mathcal{T}_{\geq t}, \mathcal{T}_{>t}$.

Definition 14. A **causal structure** is simply a DAG with the extra classification of each node into one of two categories; the **latent nodes** and **observed nodes** denoted $\mathcal{N}_L$ and $\mathcal{N}_O$. The latent nodes correspond to random variables that are either hidden through some fundamental process or cannot/will not be measured. The observed nodes are random variables that are measurable. Every node is either latent or observed and no node is both:

$$\mathcal{N}_L \cap \mathcal{N}_O = \emptyset \qquad \mathcal{N}_L \cup \mathcal{N}_O = \mathcal{N}$$

Definition 15. The **product distribution** two distributions is denoted as usual with $\times$ and is defined as,

$$(P_v \times P_w)\,(o[v]\,, o[w]) \equiv P_v(o[v])\, P_w(o[w])$$

A product distribution of k distributions is defined recursively,

$$\prod_{i=1}^{k} P_{v_i} \equiv (P_{v_1} \times \cdots \times P_{v_k})$$

Definition 16. The **marginalization** of a distribution $P_{v\cup w}$ to the distribution P_v is denoted $\sum_w P_{v\cup w} = P_v$ and is defined such that,

$$\forall o[v] \in O_v : \left(\sum_w P_{v,w}\right)(o[v]) \equiv \sum_{o[w]\in O_w} P_{v,w}(o[v]\,, o[w])$$

Todo (TC Fraser): How many definitions do I need to write??

III. TRIANGLE SCENARIO

Todo (TC Fraser): Discuss the triangle scenario, previous work done on it, etc. Focusing on the inflation depicted in fig. 2, we obtained the maximally pre-injectable sets through the procedure outlined in [2].

$$\begin{array}{cc}
\textbf{Maximal Pre-injectable Sets } \Pi & \textbf{Ancestral Independences} \\
\{A_1, B_1, C_1, A_4, B_4, C_4\} & \{A_1, B_1, C_1\} \perp \{A_4, B_4, C_4\} \\
\{A_1, B_2, C_3, A_4, B_3, C_2\} & \{A_1, B_2, C_3\} \perp \{A_4, B_3, C_2\} \\
\{A_2, B_3, C_1, A_3, B_2, C_4\} & \{A_2, B_3, C_1\} \perp \{A_3, B_2, C_4\} \\
\{A_2, B_4, C_3, A_3, B_1, C_2\} & \{A_2, B_4, C_3\} \perp \{A_3, B_1, C_2\} \\
\{A_1, B_3, C_4\} & \{A_1\} \perp \{B_3\} \perp \{C_4\} \\
\{A_1, B_4, C_2\} & \{A_1\} \perp \{B_4\} \perp \{C_2\} \\
\{A_2, B_1, C_4\} & \{A_2\} \perp \{B_1\} \perp \{C_4\} \\
\{A_2, B_2, C_2\} & \{A_2\} \perp \{B_2\} \perp \{C_2\} \\
\{A_3, B_3, C_3\} & \{A_3\} \perp \{B_3\} \perp \{C_3\} \\
\{A_3, B_4, C_1\} & \{A_3\} \perp \{B_4\} \perp \{C_1\} \\
\{A_4, B_1, C_3\} & \{A_4\} \perp \{B_1\} \perp \{C_3\} \\
\{A_4, B_2, C_1\} & \{A_4\} \perp \{B_2\} \perp \{C_1\}
\end{array} \tag{2}$$

As can be counted, there are 12 maximally pre-injectable sets which will be indexed 1 through 12 in the order seen above ($\Pi = \{\Pi_1, \dots, \Pi_{12}\}$)

IV. SUMMARY OF THE INFLATION TECHNIQUE

The causal inflation technique, first pioneered by Wolfe, Spekkens, and Fritz [2] and inspired by the *do calculus* and *twin networks* of Ref. [3], is a family of causal inference techniques that can be used to determine if a probability distribution is compatible or incompatible with a given causal structure. As a preliminary summary, the inflation technique begins by *augmenting* a causal structure with additional nodes, producing the *inflated* causal structure, and then exposes how causal inference tasks on the inflated causal structure can be used to make inferences on the original causal structure. Equipped with the common graph-theoretic terminology and notation of definition 7, an inflation can be formally defined as follows:

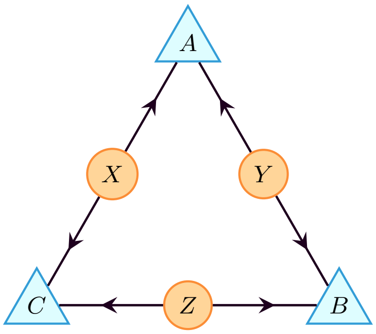

FIG. 1. The casual structure of the triangle scenario. Three variables A, B, C are observable and illustrated as triangles, while X, Y, Z are latent variables illustrated as circles.

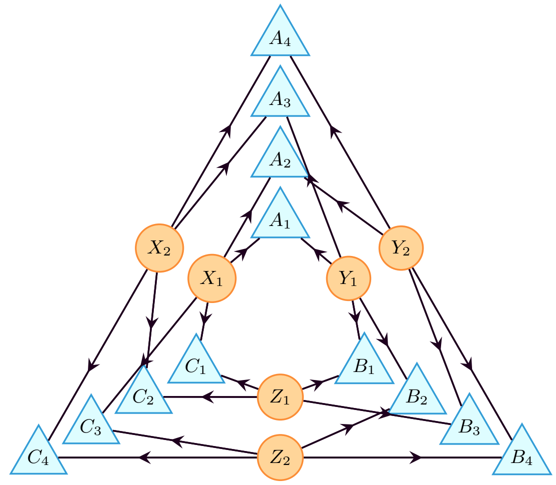

FIG. 2. An inflated causal structure of the triangle scenario fig. 1.

Definition 17. An **inflation** of a causal structure $\mathcal{G}$ is another causal structure $\mathcal{G}'$ such that:

$$\forall n' \in \mathcal{N}' : \mathrm{AnSub}_{\mathcal{G}'}(n') \sim \mathrm{AnSub}_{\mathcal{G}}(n)$$

Where '$\sim$' is notation for equivalence up to removal of the copy-index. To clarify, each node in an inflated causal structure $n' \in \mathcal{N}'$ shares a *label* assigned to a node $n \in \mathcal{N}$ in the original causal structure together with an additional index called the **copy-index**.

Definition 18. A set of **causal parameters** for a particular causal structure $\mathcal{G}$ is the specification of a conditional distribution for every node $n \in \mathcal{N}$ given it's parents in $\mathcal{G}$.

$$\left\{ P_{n|\mathrm{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N} \right\}$$

Todo (TC Fraser): Clean up what is meant by copy index, example maybe? Todo (TC Fraser): Define injectable sets Todo (TC Fraser): Define pre-injectable sets and then it's connection to probabilities Todo (TC Fraser): Define pre-injectable sets Todo (TC Fraser): State the main Compatibility lemma of inflation

V. COMPATIBILITY, CONTEXTUALITY AND THE MARGINAL PROBLEM

In order to determine if a given marginal distribution P_V or set of marginal distributions $\{P_{V_1}, \ldots, P_{V_k}\}$ is compatible with a causal structure $\mathcal{G}$, one should first formalize what is meant by *compatible*.

Definition 19. A marginal distribution P_V is **compatible** with a causal structure $\mathcal{G}$ (where it is assumed that $V \subseteq \mathcal{N}_O$) if there exists a *choice* of causal parameters $\left\{ P_{n|\mathrm{Pa}_{\mathcal{G}}(n)} \mid n \in \mathcal{N} \right\}$ such that P_V can be *recovered* from the following series of operations:
Todo (TC Fraser): Define this notation here

1. First obtain a joint distribution over *all* nodes of of the causal structure,

$$P_{\mathcal{N}} = \prod_{n \in \mathcal{N}} P_{n|\mathrm{Pa}_{\mathcal{G}}(n)}$$

2. Then marginalize over the latent nodes of $\mathcal{G}$,

$$P_{\mathcal{N}_O} = \sum_{\mathcal{N}_L} P_{\mathcal{N}}$$

3. Finally marginalize over the observed nodes not in V to obtain P_V,

$$P_V = \sum_{\mathcal{N}_O \setminus V} P_{\mathcal{N}_O}$$

A set of marginal distributions $\{P_{V_1}, \dots, P_{V_k}\}$ is compatible with $\mathcal{G}$ if each of the distributions can be made compatible by the *same* choice of causal parameters. A distribution P_V or set of distributions $\{P_{V_1}, \dots, P_{V_k}\}$ is said to be **incompatible** with a causal structure if there *does not exist* a set of causal parameters with the above mentioned property.

Todo (TC Fraser): Source this?

Operations 2 and 3 of definition 19 are related to the *marginal problem*.

Definition 20. **The Marginal Problem:** Given a set of distributions $\{P_{V_1}, \dots, P_{V_k}\}$ where $V_i \subseteq \mathcal{V}$ for some set of random variables $\mathcal{V}$ and $k \geq 2$, does there exist a joint distribution $P_{\mathcal{V}}$ such that each given distribution P_{V_i} can be obtained from marginalizing $P_{\mathcal{V}}$?

$$\forall i \in \{1, \dots, k\} : P_{V_i} = \sum_{\mathcal{V} \setminus V_i} P_{\mathcal{V}}$$

Typically (although not strictly necessary), $\mathcal{V}$ is taken to mean the union of all V_i's.

$$\mathcal{V} = V_1 \cup \dots \cup V_k = \bigcup_{i=1}^{k} V_i \tag{3}$$

Definition 21. A reoccurring motif of these discussions will be the set of distributions $\{P_{V_1}, \dots, P_{V_k}\}$ mentioned in definition 21. In agreement with [4] we will call this set of distributions a **marginal model** and denote it $P^{\mathcal{M}}$ provided that they are *compatible*:

$$\forall i \neq j \text{ if } V_i \cap V_j \neq \emptyset \text{ then } \sum_{V_i \setminus V_j} P_{V_i} = \sum_{V_j \setminus V_i} P_{V_j}$$

We call the set of subsets $\{V_1, \dots, V_k\}$ the marginal contexts or the **maximal marginal scenario** and an individual V_i a **marginal context**. Finally we will denote the union of all contexts $\mathcal{V}$ and define it exactly as in eq. (3)

Todo (TC Fraser): Discuss Compatibility, connection to cooperative games/resources, bell incompatibility? Todo (TC Fraser): Connection between contextuality and Compatibility via the marginal problem for causal parameters Todo (TC Fraser): Discuss what is meant by a 'complete' solution to the marginal problem Todo (TC Fraser): Maybe define the possibilistic marginal problem for later

VI. THE FRITZ DISTRIBUTION

The **Fritz distribution** P_F is a quantum-accessible distribution known to be incompatible with the triangle scenario. Explicitly, P_F is a three-party (A, B, C), four-outcome $(1, 2, 3, 4)$ distribution that has form as follows:

$$P_F(111) = P_F(221) = P_F(412) = P_F(322) = P_F(233) = P_F(143) = P_F(344) = P_F(434) = \frac{1}{32}\left(2 + \sqrt{2}\right)$$
$$P_F(121) = P_F(211) = P_F(422) = P_F(312) = P_F(243) = P_F(133) = P_F(334) = P_F(444) = \frac{1}{32}\left(2 - \sqrt{2}\right)$$

Here the notation $P_F(abc) = P_{ABC}(abc) = P(A = a, B = b, C = c)$ is used. The Fritz distribution P_F can be realized with the following quantum configuration:

$$\rho_{AB} = \left|\Psi^+\right\rangle\left\langle\Psi^+\right| \quad \rho_{BC} = \rho_{CA} = \left|\Phi^+\right\rangle\left\langle\Phi^+\right|$$
$$M_A = \left\{\left|0\psi_0\right\rangle\left\langle 0\psi_0\right|, \left|0\psi_\pi\right\rangle\left\langle 0\psi_\pi\right|, \left|1\psi_{-\pi/2}\right\rangle\left\langle 1\psi_{-\pi/2}\right|, \left|1\psi_{\pi/2}\right\rangle\left\langle 1\psi_{\pi/2}\right|\right\}$$
$$M_B = \left\{\left|\psi_{\pi/4}0\right\rangle\left\langle\psi_{\pi/4}0\right|, \left|\psi_{5\pi/4}0\right\rangle\left\langle\psi_{5\pi/4}0\right|, \left|\psi_{3\pi/4}1\right\rangle\left\langle\psi_{3\pi/4}1\right|, \left|\psi_{-\pi/4}1\right\rangle\left\langle\psi_{-\pi/4}1\right|\right\}$$
$$M_C = \{|00\rangle\langle 00|, |01\rangle\langle 01|, |10\rangle\langle 10|, |11\rangle\langle 11|\}$$

Where for convenience of notation ψ_x is used to denote the superposition,

$$|\psi_x\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{ix}|1\rangle\right)$$

Additionally $|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ and $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ are two maximally entangled Bell states. Fritz first proved it's incompatibility [5] by showing C acts a moderator to ensure measurement pseudo-settings for A and B are independent, satisfying non-broadcasting requirements for the standard Bell scenario. In fact, by coarse-graining outcomes for A and B and treating C as a measurement-setting moderator, P_F maximally violates the CHSH inequality. To illustrate this, begin with the CHSH inequality [6],

$$\langle AB|S_A = 1, S_B = 1\rangle + \langle AB|S_A = 1, S_B = 2\rangle + \langle AB|S_A = 2, S_B = 1\rangle - \langle AB|S_A = 2, S_B = 2\rangle \leq 2 \tag{4}$$

Where $\langle AB|S_A = i, S_B = j\rangle$ is the correlation between A and B given the measurement settings for A (B) is i (j) respectively. Next, each of C's outcomes become the condition settings in eq. (4),

$$\langle AB|C = 2\rangle + \langle AB|C = 3\rangle + \langle AB|C = 4\rangle - \langle AB|C = 1\rangle \leq 2$$

Finally, specifying the correlation between A and B to be defined in terms of a $\{1, 2, 3, 4\} \to \{(1, 4), (2, 3)\}$ coarse-graining,

$$\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} - \frac{-\sqrt{2}}{2} \leq 2$$

$$2\sqrt{2} \leq 2$$

Which corresponds to the maximum quantum violation of the CHSH inequality eq. (4)

Todo (TC Fraser): Discuss non-uniqueness and relabeling Todo (TC Fraser): Summarize Problem 2.17 in fritz BBT, make it more formal

VII. CERTIFICATE INEQUALITIES

A. Casting the Inflated Marginal Problem as a Linear Program

After obtaining the maximal pre-injectable sets associated with a particular inflation, one can write the marginal problem of definition 20 as a linear program. The key observation is that marginalization is a *linear* operator that can be performed via a matrix multiplication. To do this, we will define the *marginalization matrix*.

Definition 22. The **marginalization matrix** M for a marginal scenario $\{V_1, \dots, V_k\}$ is a bit-wise matrix where the columns are indexed by *joint* outcomes $o[\mathcal{V}] \in O_{\mathcal{V}}$ and the rows are indexed by *marginal* outcomes corresponding to all outcomes $o[V_i] \in O_{V_i}$ of the marginal contexts V_i. The entries of M are populated whenever a row index is extendable to a column index.

$$M_{(o[V_i], o[\mathcal{V}])} = \begin{cases} 1 & \exists o[\mathcal{V} \setminus V_i] \text{ such that } o[V_i] \times o[\mathcal{V} \setminus V_i] = o[\mathcal{V}] \\ 0 & \text{otherwise} \end{cases}$$

The marginalization matrix has $|O_{\mathcal{V}}|$ columns and $\sum_{i=1}^{k} |O_{V_i}|$ rows. The number of non-zero entries of M is a simple expression,

$$\sum_{i=1}^{k} |O_{V_i}| \left|O_{\mathcal{V} \setminus V_i}\right| = \sum_{i=1}^{k} |O_{\mathcal{V}}| = k\,|O_{\mathcal{V}}|$$

Each of the k elements of $\{V_1, \dots, V_k\}$ contributes a single non-zero entry to each column of M, resulting in $k\,|O_{\mathcal{V}}|$ total non-zero entries.

Note that the row and column indices of the marginalization matrix will be referred to very frequently. We will refer to the

Todo (TC Fraser): Computationally Efficient generation?

To illustrate this concretely, consider the following example:

Example 23. Suppose one has 4 binary random variables $\mathcal{V} = \{a, b, c\}$ in mind and 2 subsets $\{\{a, c\}, \{b\}\}$. Then the marginalization matrix is:

$$M = \begin{array}{c} (a,b,c)= \\ (a{=}0,c{=}0) \\ (a{=}0,c{=}1) \\ (a{=}1,c{=}0) \\ (a{=}1,c{=}1) \\ (b{=}0) \\ (b{=}1) \end{array} \begin{array}{c} \begin{array}{cccccccc} (0,0,0) & (0,0,1) & (0,1,0) & (0,1,1) & (1,0,0) & (1,0,1) & (1,1,0) & (1,1,1) \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

In order to describe how marginalization can be written as matrix multiplication $M \cdot x = b$, we need to describe how to define two more quantities:

Definition 24. The **joint distribution vector** $\mathcal{P}_{\mathcal{V}}$ for a probability distribution $P_{\mathcal{V}}$ is the vector whose entries are the positive, real-valued probabilities that $P_{\mathcal{V}}$ assigns to each outcome of $o[\mathcal{V}]$ of $O_{\mathcal{V}}$. $\mathcal{P}_{\mathcal{V}}$ shares the same indices as the *column* indices of M.

$$\mathcal{P}_{\mathcal{V}}^{\intercal} = [P_{\mathcal{V}}(o[\mathcal{V}])]_{o[\mathcal{V}] \in O_{\mathcal{V}}}$$

Definition 25. The **marginal distribution vector** $\mathcal{P}_{\{V_1,\dots,V_k\}}$ for a marginal model $\{P_{V_1}, \dots, P_{V_k}\}$ is the vector whose entries are probabilities over the set of **marginal outcomes** $\bigcup_{j=1}^{k} O_{V_j}$. $\mathcal{P}_{\{V_1,\dots,V_k\}}$ shares the same indices as the *row* indices of M.

$$\mathcal{P}_{\{V_1,\dots,V_k\}}^{\intercal} = [P_{V_i}(o[V_i])]_{o[V_i] \in \bigcup_{j=1}^{k} O_{V_j}}$$

The marginal and joint distribution vectors are related via the marginalization matrix M. Given a joint distribution vector $\mathcal{P}_{\mathcal{V}}$ one can obtain the marginal distribution vector $\mathcal{P}_{\{V_1,\dots,V_k\}}$ by multiplying M by $\mathcal{P}_{\mathcal{V}}$.

$$\mathcal{P}_{\{V_1,\dots,V_k\}} = M \cdot \mathcal{P}_{\mathcal{V}} \tag{5}$$

Todo (TC Fraser): Discuss non-unique but consistent ordering of M, $\mathcal{P}_{\mathcal{V}}$ and $\mathcal{P}_{\{V_1,\dots,V_k\}}$

The marginal problem can now be rephrased in the language of the marginalization matrix. Suppose one obtains a marginal distribution vector $\mathcal{P}_{\{V_1,\dots,V_k\}}$. The marginal problem becomes equivalent to the question: *Does there exist a joint distribution vector $\mathcal{P}_{\mathcal{V}}$ such that eq. (5) holds?*

Definition 26. **The Marginal Linear Program** is the following linear program:

$$\begin{aligned} \text{minimize:} \quad & \emptyset \cdot x \\ \text{subject to:} \quad & x \succeq 0 \\ & M \cdot x = \mathcal{P}_{\{V_1,\dots,V_k\}} \end{aligned}$$

If this "optimization"[1] is *feasible*, then there exists a vector x than can satisfy eq. (5) and is a valid joint distribution vector. Therefore feasibility implies that $P_{\mathcal{V}} = x$, solving the marginal problem with positive result. Moreover if the marginal linear program is *infeasible*, then there *does not* exist a joint distribution $P_{\mathcal{V}}$ over all random variables.

B. Infeasibility Certificates

The dual marginal linear program also provides an answer to the marginal problem. To prove this, first notice that the dual problem is *never infeasible*; by choosing y to be trivially the null vector $\emptyset$ of appropriate size, all constraints are satisfied. Secondly if $y \cdot M \succeq 0$ and $x \succeq 0$, then the following must hold if the primal is feasible:

$$y \cdot \mathcal{P}_{\{V_1,\dots,V_k\}} = y \cdot M \cdot x \geq 0 \tag{6}$$

[1] "Optimization" is presented in quotes here because the minimization objective is trivially always zero ($\emptyset$ denotes the null vector of all zero entries). The primal value of the linear program is of no interest, all that matters is its *feasibility*.

Therefore the *sign* of the dual value $d \equiv \min\left(y \cdot \mathcal{P}_{\{V_1,\ldots,V_k\}}\right)$ solves the marginal problem. If $d < 0$ then eq. (6) is violated and therefore the marginal problem has negative result. Likewise if d satisfies eq. (6), then a joint distribution $P_{\mathcal{V}}$ exists. Before continuing, an important observation needs to be made. If $d \geq 0$, then it is exactly $d = 0$, due to the existence of the trivial $y = \emptyset$. This observation is an instance of the *Complementary Slackness Property* of [7]. Comment (TC Fraser): Is this really the CSP? Moreover, if $d < 0$, then it is unbounded $d = -\infty$. This latter point becomes clear upon recognizing that for any y such that $d < 0$, another y' can be constructed by multiplying y by a real constant α greater than one such that,

$$y' = \alpha y \mid \alpha > 1 \implies d' = \alpha d < d$$

Since a more negative d' can always be found, it must be that d is unbounded. This is a demonstration of the fundamental *Unboundedness Property* of [7]; if the dual is unbounded, then the primal is infeasible.

---

[1] Tobias Fritz, "Beyond bell's theorem ii: Scenarios with arbitrary causal structure," (2014), 10.1007/s00220-015-2495-5, arXiv:1404.4812.

[2] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz, "The inflation technique for causal inference with latent variables," (2016).

[3] Judea Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2009).

[4] Tobias Fritz and Rafael Chaves, "Entropic inequalities and marginal problems," (2011), 10.1109/TIT.2012.2222863, arXiv:1112.4788.

[5] Tobias Fritz, "Beyond bell's theorem: Correlation scenarios," (2012), 10.1088/1367-2630/14/10/103001, arXiv:1206.5115.

[6] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, "Proposed experiment to test local hidden-variable theories," Phys. Rev. Lett. **23**, 880–884 (1969).

[7] Judea Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, 2009).