Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-34 (2019) 47–52

Non-Commutative Probability Models in Human Decision Making: Binary Hypothesis Testing ⋆

Aneesh Raghavan * **John S. Baras** *

* *Institute for Systems Research, University of Maryland, College Park, USA (e-mail: raghava,baras@umd.edu).*

Abstract: In this paper, we consider the binary hypothesis testing problem, as the simplest human decision making problem, using a von-Neumann non-commutative probability framework. We present two approaches to this decision making problem. In the first approach, we represent the available data as coming from measurements modeled via projection valued measures (PVM) and retrieve the results of the underlying detection problem solved using classical probability models. In the second approach, we represent the measurements using positive operator valued measures (POVM). We prove that the minimum probability of error achieved in the second approach is the same as in the first approach.

Keywords: Hypothesis testing, PVM, POVM

1. INTRODUCTION

Networked multi-agent systems are ubiquitous systems which have become an integral part of life and work. Examples of such systems include networked vehicles, smart rooms and, collaborative robots. Often, humans interact with these systems. Hence, it has become essential to understand how humans make decisions and judgments under uncertainty. Psychologists have studied human cognition and decision making for a long time, Tversky and Kahneman (1974), Tversky and Kahneman (1981), Frisch and Clemen (1994) and Buchanan and O Connell (2006). Recent experiments, Trueblood and Busemeyer (2012), Busemeyer and Bruza (2012), Khrennikov et al. (2014), Wang et al. (2014), Busemeyer et al. (2015), Trueblood et al. (2017), have reavealed the following characteristics of human judgments: (i) When a question is asked, the state of the human changes from an indefinite state to a definite state with respect to the question; (ii) The order in which questions are asked can influence the answers to the questions; (iii) The law of total probability is violated. For more details, we refer to chapter 1 of Busemeyer and Bruza (2012) and the other reference mentioned above. Let us consider a classical probability model for human decision making. The sample space could include events from politics, science, finance, sports, etc. Finding the joint distribution between all possible events in these domains is difficult. Hence, it is reasonable to assume that all these events do not belong to the same sample space. Further, fallacies like the conjunction and disjunction fallacy are observed when classical probability models are used to model human decision making, Busemeyer and Bruza (2012). Psychologists have used Quantum Probability Models (QPM) to explain some of the fallacies observed (Busemeyer and Bruza (2012)). That is the underlying probability models are von-Neumann probability models, Neumann (1955), where the events are represented by subspaces of a Hilbert space rather than subsets of a set (as in the classical Kolmogorov probability theory), Hayashi (2006), Holevo (2003). There is no quantum physics involved here, just the underlying logic of events is different than the classical one. These are the non-commutative probability models in the title of the paper.

In Trueblood and Busemeyer (2012), three different situations are examined, where judgments related to causal inference problems produce unexpected results. Von Neumann probability models (VNPM) were used to explain the experimental findings in all three situations. It has also been shown that other models like Bayes net and belief adjustment model only account for a subset of findings. In Trueblood et al. (2017), the authors propose a hierarchy of models for human cognition which could be adopted in different situations. The different models arise due to different assumptions imposed on the existence of joint distribution between the events. Using the results of three experiments they show that their modeling approach explains five key phenomena in human inference including the order effects, reciprocity, memorylesness, violations of Markov conditions and anti-discounting. In Khrennikov et al. (2014), the authors analyze the applicability of VNPM to two basic properties of opinion polling: (i)response replicability and (ii) question order effect. It is mentioned that VNPM can account for one of the properties but not both. They conclude that either problems where VNPM are applicable need to be characterized or more general representations than POVM are needed. In Borie (2013), the authors reformulate von Neumann and Morgenstern's approach, Von Neumann and Morgenstern (1947), to modeling human decision maker behavior using non-commutative probability theory. They

⋆ Research supported by ARO grant W911NF1410384 and by ONR grant N00014-17-1-2622, and the Knut and Alice Wallenberg Foundation, the Swedish Strategic Research Foundation and the Swedish Research Council.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2019.01.019

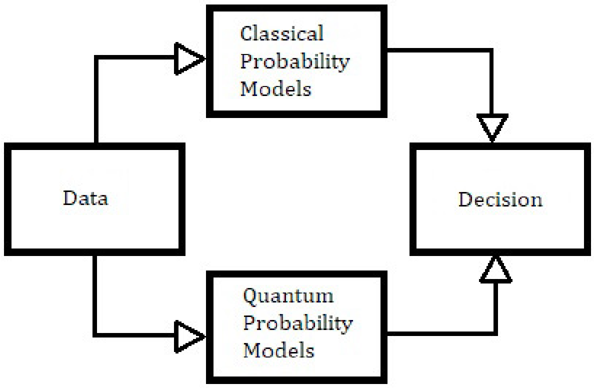

Fig. 1. Two possible models

show that their generalization allows for decision makers to make a distinction between representations of a set of events and enables several paradoxes and inconsistencies in traditional expected utility theory (e.g., Allais paradox, etc.) to be better understood. In Frisch and Clemen (1994), the authors assert that a good decision making process must be concerned with how decision makers evaluate potential consequences of decisions, the extent to which they identify and evaluate the risks associated with the decisions and their approach to making the final choices.

The question that arises is, given data which of the two probability models (classical Kolmogorov-like and non-classical von Neumann-like probability models) is more appropriate for modeling human decision policies (figure 1), measured by the relative accuracy of the model's predictions for actual human decisions. Hence two fundamental questions arise: (i) how to represent(i.e. model) measurements (and the data they generate), (ii) what metrics to use to model human decision making. In this paper, the objective is to compare the probability of error achieved when the hypothesis testing problem is solved in the classical (Kolmogorov) probability framework and in the non classical (von Neumann, quantum-like) probability framework. We consider a single observer. In the first phase, the observer collects observations and knows the true hypothesis under which the observations were generated. Using the frequentist approach empirical distributions are built. In the second phase, using the empirical distributions, hypothesis testing problems are formulated in classical probability and non-classical (non-commutative) probability frameworks. When the measurements are represented by a projection valued measure (PVM), Baras et al. (1976), Baras (1979), Baras and Harger (1977), Helstrom (1969), Holevo (2003), we show that the results of the detection problem (optimal cost and decision policy) are identical to that of the classical probability models. When measurements are incompatible, they are represented using positive operator valued measures (POVM), Baras (1987), Baras (1988), Holevo (2003). The detection problem is solved and through an example, we discuss the effect of order in measurements on the minimum probability of error.

Notation: Let $\mathcal{H}$ be a complex Hilbert space. Then $\mathcal{B}(\mathcal{H})(\mathcal{B}_s^+(\mathcal{H}))$ denotes the space of bounded (positive, Hermitian and bounded) operators which map from $\mathcal{H}$ to $\mathcal{H}$. $\mathcal{T}(\mathcal{H})(\mathcal{T}_s^+(\mathcal{H}))$ are subsets of $\mathcal{B}(\mathcal{H})(\mathcal{B}_s^+(\mathcal{H}))$ such that trace of the operators is 1. $\mathcal{P}(\mathcal{H})$ is the subset of $\mathcal{B}_s^+(\mathcal{H})$ such that the operators are orthogonal projections. Tr denotes the trace operator. For an operator $O \in \mathcal{B}(\mathcal{H})$, O^H denotes its conjugate transpose (The Hilbert spaces here are finite dimensional).

2. PROBLEM FORMULATION

We consider a single observer. The observation collected by the observer is denoted by Y, $Y \in S$, $|S| = N$ where S is a finite set of real numbers or real vectors of finite dimension. Data strings consisting of observation and true hypothesis are collected by the observer. From the data strings, empirical distributions are built. Let $p_i^h, 1 \leq i \leq N$ be the distribution under hypothesis h. The prior probabilities of hypotheses can be found from the data and are represented by ζ_1 (for $H = 1$) and ζ_0 (for $H = 0$). In the quantum probability framework, there are multiple ways in which measurements can be captured. Two of them are: (a) Projection valued measures (PVM) (b) Positive operator valued measures (POVM). In this section we discuss the formulation of the detection problem in classical probability framework and von Neumann probability framework with both representations for measurements.

2.1 Classical Probability

Let $\Omega = \{0,1\} \times S$ be the sample space. Let $\mathcal{F} = 2^\Omega$ be the associated algebra. An element in the sample space can be represented by $\omega = (h, y)$, where $h \in \{0,1\}$ and $y \in S$. The measure is $\mathbb{P}(\omega) = \zeta_h p_y^h$. The probability space is $(\Omega, \mathcal{F}, \mathbb{P})$. Given a new observation, $Y = y$ the detection problem is to find D such that the following cost is minimized:

$$\mathbb{E}_{\mathbb{P}}[H(1-D) + (1-H)D],$$

i.e, the probability of error is minimized. H represents the hypothesis random variable. Once the decision is found the optimal cost also needs to be found.

2.2 Projection valued measure

Projection Valued Measure(PVM): Let (X, Σ) be a measurable space. A projection valued measure is a mapping F from Σ on to $\mathcal{P}(\mathcal{H})$ such that

(i) $F(X) = \mathbb{I}$.
(ii) $A, B \in \Sigma$ such that $A \cap B = \emptyset$, then $F(A \cup B) = F(A) + F(B)$.
(iii) If $\{A_i\}_{i \geq 1} \subseteq \Sigma$, such that $A_1 \subset A_2 \subset ...$, then $F(\cup_{i=1}^{\infty} A_i) = \lim_{i \to \infty} F(A_i)$.

For the detection problem, $X = \{1, 2, ..., N\}$, $\Sigma = 2^X$. The second condition implies that the minimum dimension of the complex Hilbert space in consideration is N. We let $\mathcal{H} = \mathbb{C}^N$. The first objective is to find $\rho_h \in \mathcal{T}_s^+(\mathbb{C}^N)$ and $F : \Sigma \to \mathcal{P}(\mathbb{C}^N)$, such that

$$Tr[\rho_h F(i)] = p_i^h, h = 0, 1, 1 \leq i \leq N, \tag{1}$$

$$F(i)F(j) = \Theta_{\mathbb{C}^N}\ , 1 \leq i, j \leq N, i \neq j \text{ and } \sum_{i=1}^{N} F(i) = \mathbb{I}_{\mathbb{C}^N}, \tag{2}$$

where $\Theta_{\mathbb{C}^N}$ is zero operator and $\mathbb{I}_{\mathbb{C}^N}$ is identity operator. Given the state and the PVM, we consider the formulation of the detection problem mentioned in Baras (1979), section 3.4. Let C_{ij} denote the cost incurred when the decision made is i while the true hypothesis is j. Since the objective is to minimize the probability of error, we let $C_{10} = 1$, $C_{00} = 0$, $C_{01} = 1$ and $C_{11} = 0$. The decision policy, $\{\alpha_i^1\}_{i=1}^N$ and $\{\alpha_i^0\}_{i=1}^N$ denotes the probability of choosing $D = 1$ and $D = 0$ respectively when observation i is received. Given observation i, the probability of choosing $D = 1$ (D is the decision) and the true hypothesis being 0 is $\zeta_0 Tr[\rho_0 F(i)]\alpha_i^1$. Hence the probability of choosing $D = 1$ and true hypothesis being 0 is $\sum_{i=1}^N \zeta_0 Tr[\rho_0 F(i)]\alpha_i^1$. Similarly, the probability of choosing $D = 0$ and true hypothesis being 1 is $\sum_{i=1}^N \zeta_1 Tr[\rho_1 F(i)]\alpha_i^0$. Hence the probability of error is :

$$\begin{aligned}\mathbb{P}_e &= \sum_{i=1}^N \zeta_0 Tr[\rho_0 F(i)]\alpha_i^1 + \sum_{i=1}^N \zeta_1 Tr[\rho_1 F(i)]\alpha_i^0 \\ &= Tr\left[\zeta_0\rho_0\left[\sum_{i=1}^N \alpha_i^1 F(i)\right] + \zeta_1\rho_1\left[\sum_{i=1}^N \alpha_i^0 F(i)\right]\right]\end{aligned}$$

We define the risk operators as :

$$W_1 = \zeta_0\rho_0,\ W_0 = \zeta_1\rho_1$$

and note that,

$$\sum_{i=1}^N \alpha_i^h F(i) \geq 0, h = 0,1 \sum_{i=1}^N \left[\alpha_i^1 F(i) + \alpha_i^0 F(i) = \mathbb{I}_{\mathbb{C}^N}\right].$$

Instead of minimizing over the decision policies, we minimize over pairs of operators which are semi-definite and sum to identity. Hence, the detection problem is formulated as follows

$$\begin{aligned}P1 : \min_{\Pi_1,\Pi_0} \quad & Tr[W_1\Pi_1 + W_0\Pi_0] \\ \text{s.t} \quad & \Pi_1 \in \mathcal{B}_s^+(\mathbb{C}^N),\ \Pi_0 \in \mathcal{B}_s^+(\mathbb{C}^N), \\ & \Pi_1 + \Pi_0 = \mathbb{I}_{\mathbb{C}^N}\end{aligned}$$

The solution of the above problem, Π_1^*, Π_0^* are the detection operators which are to be realized using the PVM:

$$\begin{aligned}P2 : \quad & \exists\{\alpha_i^1\}_{i=1}^N \text{ and } \{\alpha_i^0\}_{i=1}^N \\ \text{s.t} \quad & \alpha_i^1 \geq 0,\ \alpha_i^0 \geq 0,\ \alpha_i^1 + \alpha_i^0 = 1,\ 1 \leq i \leq N, \\ \text{and} \quad & \Pi_1^* = \sum_{i=1}^n \alpha_i^1 F(i),\ \Pi_0^* = \sum_{i=1}^n \alpha_i^0 F(i).\end{aligned}$$

Suppose for two pairs of states, (ρ_1, ρ_0), $(\bar{\rho}_1, \bar{\rho}_0)$ and PVM F, (1) is satisfied,i.e.,

$$Tr[\rho_h F(i)] = Tr[\bar{\rho}_h F(i)] = p_i^h, h = 0, 1, 1 \leq i \leq N.$$

If we consider the solution to P1 alone, the corresponding detection operators (Π_1^*, Π_0^*), $(\bar{\Pi}_1^*, \bar{\Pi}_0^*)$ and the minimum costs achieved, $\mathbb{P}_e, \bar{\mathbb{P}}_e$ could be different. However, if we consider solution to P1 such that P2 is feasible, i.e., the detection operators are realizable, then,

$$\begin{aligned}\mathbb{P}_e &= Tr[W_1\Pi_1^* + W_0\Pi_0^*] \\ &= \sum_{i=1}^N \zeta_0 Tr[\rho_0 F(i)]\alpha_i^1 + \sum_{i=1}^N \zeta_1 Tr[\rho_1 F(i)]\alpha_i^0 \\ &= \sum_{i=1}^N \zeta_0 Tr[\bar{\rho}_0 F(i)]\alpha_i^1 + \sum_{i=1}^N \zeta_1 Tr[\bar{\rho}_1 F(i)]\alpha_i^0 \geq \bar{\mathbb{P}}_e.\end{aligned}$$

Similarly, $\bar{\mathbb{P}}_e \geq \mathbb{P}_e$. Hence $\bar{\mathbb{P}}_e = \mathbb{P}_e$. Hence for a fixed PVM representation, the optimal cost does not change with different state representations.

2.3 Positive operator valued measure

Consider the scenario where $S \subset \mathbb{R}^2$. In such a scenario the observer collects two observations, Y_1 and Y_2. Two measurements are said to be incompatible if they cannot be measured simultaneously. The joint distribution between the measurements does not exist. If Y_1 and Y_2 are incompatible, then the order in which they are measured could lead to different outcomes. Let $Y_1 \in Z_1$, $|Z_1| = \eta_1$ and $Y_2 \in Z_2, |Z_2| = \eta_2$. Then Y_1 and Y_2 can be individually represented as PVMs in Hilbert space of dimension η, $\eta = \max\{\eta_1, \eta_2\}$. Let the PVM corresponding to Y_1 and Y_2 be μ and ν respectively. Let the state be ρ. Suppose Y_1 is measured first and value obtained is $i \in Z_1$. Then the state after measurement of Y_1 changes from ρ to (Davies (1976)) :

$$\rho_i = \frac{\mu(i)\rho\mu(i)}{Tr[\rho\mu(i)]}.$$

After measuring Y_1, Y_2 is measured. The conditional probability of $Y_2 = j$ given $Y_1 = i$ is,

$$Tr[\rho_i\nu(j)] = \frac{Tr[\mu(i)\rho\mu(i)\nu(j)]}{Tr[\rho\mu(i)]} = \frac{Tr[\rho\mu(i)\nu(j)\mu(i)]}{Tr[\rho\mu(i)]}.$$

Thus the probability of obtaining $Y_1 = i$ and then $Y_2 = j$ is $Tr[\rho\mu(i)\nu(j)\mu(i)]$. Further, the measurement corresponding to Y is, $\sigma_1(i,j) = \mu(i)\nu(j)\mu(i), 1 \leq i \leq \eta_1, 1 \leq j \leq \eta_2$. If Y_1 is measured after Y_2, then the measurement corresponding to Y is, $\sigma_2(i,j) = \nu(i)\mu(j)\nu(i), 1 \leq i \leq \eta_2, 1 \leq j \leq \eta_1$. Since for any (i,j), $\mu(i)$ and $\nu(j)$ do not commute, $\sigma_1(i,j)$ and $\sigma_2(i,j)$ are not projections. They are positive, Hermitian and bounded. Hence σ_1, σ_2 are not PVMs, and belong to a larger class of measurements, i.e., the POVMs.

Positive Operator Valued Measure (POVM): Let (X, Σ) be a measurable space. A positive operator valued measure is a mapping M from Σ on to $\mathcal{B}_s^+(\mathcal{H})$ such that, if $\{X_i\}_{i\geq 1}$ is partition of X, then

$$\sum_i M(X_i) = \mathbb{I} \text{ (Strong Topology)}$$

Further for $A, B \in \Sigma$ such that $A \cap B = \emptyset$, if $M(A)M(B) = \Theta_{\mathcal{H}}$, then M is a PVM. We consider the dimension of the Hilbert space to be k, $k \geq 2$. As in the previous formulation, the first objective is to find states, $\hat{\rho}_h \in \mathcal{T}_s^+(\mathbb{C}^k), h = 0,1$ and POVM, $M : \Sigma \to \mathcal{B}_s^+(\mathbb{C}^k)$ such that

$$Tr[\hat{\rho}_h M(i)] = p_i^h, h = 0, 1, 1 \leq i \leq N \text{ and} \sum_{i=1}^N M(i) = \mathbb{I}_{\mathbb{C}^k}. \tag{3}$$

The probability of error calculation is analogous to the previous section. We define the new risk operators as:

$$\hat{W}_1 = \zeta_0\hat{\rho}_0,\ \hat{W}_0 = \zeta_1\hat{\rho}_1$$

Given states and POVM, the detection problem with the same cost parameters as P1, is formulated as:

$$\begin{aligned}P3 : \min_{\hat{\Pi}_1,\hat{\Pi}_0} \quad & Tr[\hat{W}_1\hat{\Pi}_1 + \hat{W}_0\hat{\Pi}_0] \\ \text{s.t} \quad & \hat{\Pi}_1 \in \mathcal{B}_s^+(\mathbb{C}^k),\ \hat{\Pi}_0 \in \mathcal{B}_s^+(\mathbb{C}^k) \\ & \hat{\Pi}_1 + \hat{\Pi}_0 = \mathbb{I}_{\mathbb{C}^k}.\end{aligned}$$

The decision policies $\{\beta_i^1\}_{i=1}^N$ and $\{\beta_i^0\}_{i=1}^N$ are found by solving the following problem:

$$
\begin{aligned}
P4: \quad & \exists\{\beta_i^1\}_{i=1}^N \text{ and } \{\beta_i^0\}_{i=1}^N \\
\text{s.t} \quad & \beta_i^1 \geq 0,\ \beta_i^0 \geq 0,\ \beta_i^1 + \beta_i^0 = 1,\ 1 \leq i \leq N, \\
\text{and} \quad & \hat{\Pi}_1 = \sum_{i=1}^n \beta_i^1 M(i),\ \hat{\Pi}_0 = \sum_{i=1}^n \beta_i^0 M(i).
\end{aligned}
$$

Consider the problem:

$$
\begin{aligned}
P5: \quad & \min_{\hat{\Pi}_1, \hat{\Pi}_0, \{\beta_i^1\}_{i=1}^N} Tr[\hat{W}_1\hat{\Pi}_1 + \hat{W}_0\hat{\Pi}_0] \\
\text{s.t} \quad & \hat{\Pi}_1 \in \mathcal{B}_s^+(\mathbb{C}^k),\ \hat{\Pi}_0 \in \mathcal{B}_s^+(\mathbb{C}^k) \\
& \hat{\Pi}_1 + \hat{\Pi}_0 = \mathbb{I}_{\mathbb{C}^k}. \\
& 0 \leq \beta_i^1 \leq 1, 1 \leq i \leq n. \\
& \hat{\Pi}_1 = \sum_{i=1}^n \beta_i^1 M(i),\ \hat{\Pi}_0 = \sum_{i=1}^n (1-\beta_i^1) M(i).
\end{aligned}
$$

Let the feasible set of detection operators for $P3$ be S_1 and for $P5$ be S_2. Due to additional constraints in $P5$, $S_2 \subseteq S_1$. The detection operators obtained by solving $P3$ may or may not be realizable,i.e., $P4$ may not be feasible. In $P5$, the optimization is only over detection operators which are realizable. If the solution of $P3$ is such that $P4$ is feasible then it is the solution for $P5$ as well. It is also possible that $P3$ is solved, $P4$ is not feasible and $P5$ is solved. The objective is to understand the minimum probability of error which can achieved by detection operators which are realizable. Hence, we consider the solution of $P5$ and compare it with the minimum error achieved in PVM approach.

Let $\mathbb{M}$ be set of all POVMs on Σ. Let $\mathbb{S} \subset \mathcal{T}_s^+(\mathbb{C}^k) \times \mathcal{T}_s^+(\mathbb{C}^k) \times \mathbb{M}$ be the set of, pairs of states and a POVM such that (3) is satisfied. Let $\bar{\mathbb{S}} \subseteq \mathbb{S}$ be the triples for which the optimization problem $P5$ can be solved. For a triple $(\hat{\rho}_0, \hat{\rho}_1, M)$ in $\bar{\mathbb{S}}$, we define $Q(\hat{\rho}_0, \hat{\rho}_1, M)$ to be the optimal value achieved by solving $P5$.

3. SOLUTION

3.1 Classical Probability

It suffices to minimize,

$$\mathbb{E}_{\mathbb{P}}[H(1-D) + (1-H)D | Y = y].$$

$\mathbb{E}_{\mathbb{P}}[H|Y = y] = \frac{p_y^1\zeta_1}{p_y^1\zeta_1 + p_y^0\zeta_0}$. $D = 1$ if $p_y^1\zeta_1 \geq p_y^0\zeta_0$ else $D = 0$. Thus the cost paid when the observation is y is $\frac{\min\{p_y^1\zeta_1, p_y^0\zeta_0\}}{p_y^1\zeta_1 + p_y^0\zeta_0}$. The expected cost is :

$$\sum_{i=1}^N \left[\frac{\min\{p_i^1\zeta_1, p_i^0\zeta_0\}}{p_i^1\zeta_1 + p_i^0\zeta_0}\right] \times \mathbb{P}(Y = i) = \sum_{i=1}^N \min\{p_i^1\zeta_1, p_i^0\zeta_0\}.$$

3.2 Projection valued measure

Define,

$$\rho_h = \begin{bmatrix} p_1^h & & \\ & \ddots & \\ & & p_N^h \end{bmatrix} \text{ and } F(i) = e_i e_i^H,$$

where e_i represents the canonical basis in $\mathbb{C}^N$. Clearly equations (1) and (2) are satisfied.

Theorem 1. (Baras (1979), Baras (1987)). There exists a solution to the problem

$$\min Tr[W_0\Pi_0 + W_1\Pi_1]$$

over all two component POM's , where $W_0, W_1 \in \mathcal{B}_s^+(\mathbb{C}^N)$. A necessary and sufficient condition for Π_i^* to be optimal is that:

$$W_0\Pi_0^* + W_1\Pi_1^* \leq W_i, i = 0, 1 \tag{4}$$
$$\Pi_0^*W_0 + \Pi_1^*W_1 \leq W_i, i = 0, 1 \tag{5}$$

Furthermore, under any of above conditions the operator

$$O = W_0\Pi_0^* + W_1\Pi_1^* = \Pi_0^*W_0 + \Pi_1^*W_1$$

is self-adjoint and unique solution to the dual problem.

To solve P1, we invoke the above theorem. Hence Π_1^* and Π_0^* are optimal for $P1$ if and only if

$$W_1\Pi_1^* + W_0\Pi_0^* \leq W_1,\ W_1\Pi_1^* + W_0\Pi_0^* \leq W_0$$
$$\Pi_1^*, \Pi_0^* \in\ \mathcal{B}^+(\mathbb{C}^N) \text{and } \Pi_1^* + \Pi_0^* = \mathbb{I}_{\mathbb{C}^N}.$$

Further, Π_1^* and Π_0^* are optimal for $P1$ if and only if they satisfy the above constraints and are diagonal matrices. The realisability condition in $P2$ forces Π_1^* and Π_0^* to be diagonal matrices. Let $\Pi_1^* = \text{diag}(n_1^1, \ldots, n_N^1)$ and $\Pi_0^* = \text{diag}(1-n_1^1, \ldots, 1-n_N^1)$. Then for optimality,

$$\text{for } 1 \leq i \leq N, \begin{cases} \zeta_0 p_i^0 n_i^1 + \zeta_1 p_i^1(1-n_i^1) \leq \zeta_0 p_i^0, \\ \zeta_0 p_i^0 n_i^1 + \zeta_1 p_i^1(1-n_i^1) \leq \zeta_1 p_i^1 \end{cases}$$

For both inequalities to hold, it follows that if $\zeta_0 p_i^0 \geq \zeta_1 p_i^1$, then $n_i^1 = 0$. Else $n_i^1 = 1$. The minimum cost achieved is :

$$\mathbb{P}_e^* = \sum_{i=1}^N \min\{\zeta_0 p_i^0, \zeta_1 p_i^1\} \leq \min\{\zeta_0, \zeta_1\}.$$

Clearly $\alpha_i^j = n_i^j$, $1 \leq i \leq N$, $,j = 1, 0$. As in the classical probability scenario, we obtain pure strategies, i.e , when measurement i is obtained , if $\zeta_0 p_i^0 \geq \zeta_1 p_i^1$ then the decision is 0 with probability 1, else decision is 1 with probability 1.

Let $\bar{\rho}_h, h = 0, 1$ be another pair of states and $G : \Sigma \to \mathcal{P}(\mathbb{C}^N)$, be another PVM such that equations(1) and (2) are satisfied. Since each $G(i)$ is a rank one matrix,

$$\exists v_i \in \mathbb{C}^N \text{ s.t } v_i^H v_i = 1, G(i) = v_i v_i^H, 1 \leq i \leq n$$
$$v_i^H v_j = 0, 1 \leq i, j \leq n,\ i \neq j.$$

Let $T = [v_1; v_2, ..., v_n]$. T is a $n \times n$ matrix with its columns composed by vectors v_i. Thus,

$$T^H T = TT^H = \mathbb{I}_{\mathbb{C}^N}, T^H G(i) T = F(i), 1 \leq i \leq n$$

Since T is an isometry, $\tilde{\rho}_h = T^H \bar{\rho}_h T \in \mathcal{T}_s^+(\mathbb{C}^N), h = 0, 1$. Hence,

$$Tr[\bar{\rho}_h G(i)] = Tr[\bar{\rho}_h TT^H G(i) TT^H] =$$
$$Tr[T^H \bar{\rho}_h TT^H G(i) T] = Tr[\tilde{\rho}_h F(i)].$$

Hence the optimal cost does not change with different PVM and state representations. The proof can be extended, for state and PVM representations in $\mathbb{C}^M, M > N$.

3.3 Positive operator valued measure

To find the states and the POVM, a new numerical method is proposed. If a feasibility problem is formulated with the state and POVM as optimization variables, the resulting problem is nonconvex. Hence we consider a finite set of states, $\mathcal{S} \subset \mathcal{T}_s^+(\mathbb{C}^k)$, $|\mathcal{S}| < \infty$. For a pair of states, $(\hat{\rho}_0, \hat{\rho}_1) \in \mathcal{S} \times \mathcal{S}$, $\hat{\rho}_0 \neq \hat{\rho}_1$, the following feasibility solved:

$$P6: \quad \min_{t \in \mathbb{R}, \{M(i)\}_{i=1}^N \subset \mathbb{C}^{k\times k}} t$$

$$\text{s.t} \quad Tr[\hat{\rho}_h M(i)] - p_i^h = t, h = 0, 1, 1 \leq i \leq N$$

$$M(i) \leq -t, 1 \leq i \leq N, \sum_{i=1}^N M(i) - I_{\mathbb{C}^k} = tI_{\mathbb{C}^k}.$$

If for a particular pair of states, $\hat{\rho}_0, \hat{\rho}_1$ the optimal value of the above feasibility problem, t^* is less than or equal to zero, then the corresponding minimizers, $\{M(i)\}_{i=1}^N$ is the POVM. If for every pair of states, the optimal value of the feasibility problem is grater than zero, then optimization problems need to be solved for a new set of states.

Theorem 2. (*Naimark's dilation Theorem*), [Busch et al. (1997)]. Let $M : \Sigma \to \mathcal{B}_s^+(\mathcal{H})$ be POVM. There exists a Hilbert Space $\mathcal{K}$, a PVM $P : \Sigma \to \mathcal{P}(\mathcal{K})$ and an isometry $T : \mathcal{H} \to \mathcal{K}$ such that

$$M(S) = T^* P(S) T \ \ \forall \, S \in \Sigma,$$

where T^* is the adjoint of the operator T.

For completeness, we find the isometry T when $\mathcal{H} = \mathbb{C}^k$. For any vector $x \in \mathcal{H}$, let x_e be representation of the vector in the standard canonical basis of $\mathcal{H}$. Let $L = \bigoplus_{i=1}^N \mathcal{H}$. Let $\{e_i\}_{i=1}^{N\times k}$ be the canonical basis of L. For vector $v \in L$, there exist unique coefficients v_{ij} such that $v = \sum_{i=1}^N \sum_{j=1}^k v_{ij} e_{(i-1)\times N + j}$. Let $v_i = [v_{i1}; v_{i2}; \ldots, v_{ik}]$ and $v_e = [v_1, \ldots, v_N]$. Let $\overline{M} = \text{diag}(M(1), \ldots, M(N))$. Note that $\overline{M} = \overline{M}^H$. The inner product on L is defined as:

$$\langle v, u\rangle = v_e^H \overline{M} u_e = \sum_{i=1}^N v_i^H M(i) u_i$$

Let $\mathcal{N} = \{v \in L : \langle v, v\rangle = 0\}$. We define $\mathcal{K} = \overline{\bigoplus_{i=1}^N \mathcal{H}/\mathcal{N}}$. Thus $T : \mathcal{H} \to \mathcal{K}$ can be defined as: $T(v) = (v, \ldots, v)$. In the standard canonical basis, the matrix representation of T would be

$$V^H = [I_{\mathbb{C}^k} \ I_{\mathbb{C}^k} \ \ldots \ I_{\mathbb{C}^k},]_{k\times(N\times k)}.$$

Let the matrix representation of T^* in the canonical basis be U. From the adjoint equation it follows that $(Uy_e)^H x_e = y_e^H \overline{M} V x_e$, $\forall \ x_e \in \mathbb{C}^k$ and $\forall \ y_e \in \mathbb{C}^{N\times k}$. Hence $U = V^H \overline{M} =$

$$[M_1 \ M_2 \ \ldots \ M_N]_{k\times(N\times k)}, UV = I_{\mathbb{C}^k}.$$

Let $P : \Sigma \to \mathcal{P}(\mathbb{C}^{N\times k})$ be defined as :

$$P(i) = \begin{bmatrix} \Theta_{\mathbb{C}^k} & \Theta_{\mathbb{C}^k} & \ldots \\ \Theta_{\mathbb{C}^k} & \ldots & \\ \vdots & (I_{\mathbb{C}^k})_{i,i} & \ldots \\ & & \Theta_{\mathbb{C}^k} \end{bmatrix}_{(N\times k)\times(N\times k)}$$

$P(i)$ is a collection of N^2, $k \times k$ matrices, where the i diagonal matrix is the identity matrix and the rest are zero matrices. Hence $M(i) = UP(i)V$. Let $\tilde{\rho}_h \in \mathcal{T}_s^+(\mathbb{C}^{N\times k})$ be equal to $V\hat{\rho}_h U$ for $h = 0, 1$, then

$$\sum_{i=(j-1)\times k+1}^{j\times k} e_i^H \tilde{\rho}_h e_i = p_j^h, h = 0, 1, j = 1, 2, \ldots N.$$

Lemma 3. If $\bar{\mathbb{S}} \neq \emptyset$, let,

$$\mathbb{Q}_e^* = \min_{(\hat{\rho}_0, \hat{\rho}_1, M)\in\bar{\mathbb{S}}} Q(\hat{\rho}_0, \hat{\rho}_1, M).$$

Then,

$$\mathbb{Q}_e^* = \mathbb{P}_e^* \tag{6}$$

Proof. For a triple $(\hat{\rho}_0, \hat{\rho}_1, M) \in \bar{\mathbb{S}}$, let $(\hat{\Pi}_1^*, \hat{\Pi}_0^*)$ and $\{\beta_i^{1,*}, \beta_i^{0,*}\}_{i=1}^{i=n}$ solve P5. Then,

$$\begin{aligned} &Tr[\hat{W}_1\hat{\Pi}_1^* + \hat{W}_0\hat{\Pi}_0^*] = \\ &= Tr[\hat{W}_1 \sum_{i=1}^n \beta_i^{1,*} M(i) + \hat{W}_0 \sum_{i=1}^n \beta_i^{0,*} M(i)] \\ &= \sum_{i=1}^n \zeta_0 Tr[\hat{\rho}_0 M(i)]\beta_i^{1,*} + \zeta_1 Tr[\hat{\rho}_1 M(i)]\beta_i^{0,*} \\ &= \sum_{i=1}^n \zeta_0 Tr[\hat{\rho}_0 UP(i)V]\beta_i^{1,*} + \zeta_1 Tr[\hat{\rho}_1 UP(i)V]\beta_i^{0,*} \\ &= \sum_{i=1}^n \zeta_0 Tr[V\hat{\rho}_0 UP(i)]\beta_i^{1,*} + \zeta_1 Tr[V\hat{\rho}_1 UP(i)]\beta_i^{0,*} \\ &= \sum_{i=1}^n \zeta_0 Tr[\tilde{\rho}_0 P(i)]\beta_i^{1,*} + \zeta_1 Tr[\tilde{\rho}_1 P(i)]\beta_i^{0,*} \\ &= \sum_{i=1}^n \zeta_0 p_i^0 \beta_i^{1,*} + \zeta_1 p_i^1 (1 - \beta_i^{1,*}) \end{aligned}$$

For any other pair of realizable detection operators $(\hat{\Pi}_1, \hat{\Pi}_0)$, with decision policy $\{\beta_i^1, \beta_i^0\}_{i=1}^{i=n}$,

$$Tr[\hat{W}_1\hat{\Pi}_1^* + \hat{W}_0\hat{\Pi}_0^*] = \sum_{i=1}^n \zeta_0 p_i^0 \beta_i^1 + \zeta_1 p_i^1 (1 - \beta_i^1).$$

Hence for any decision policy $\{\beta_i^1, \beta_i^0\}_{i=1}^{i=n}$,

$$\sum_{i=1}^n \zeta_0 p_i^0 \beta_i^{1,*} + \zeta_1 p_i^1 (1 - \beta_i^{1,*}) \leq \sum_{i=1}^n \zeta_0 p_i^0 \beta_i^1 + \zeta_1 p_i^1 (1 - \beta_i^1).$$

Thus,

$$\beta_i^{1,*} = \begin{cases} 1, \text{ if, } \zeta_1 p_i^1 \geq \zeta_0 p_i^0, \\ 0, \text{ otherwise.} \end{cases}$$

$$Tr[\hat{W}_1\hat{\Pi}_1^* + \hat{W}_0\hat{\Pi}_0^*] = \sum_{i=1}^N \min\{\zeta_0 p_i^0, \zeta_1 p_i^1\} = \mathbb{P}_e^*$$

Since the above result is true for every triple in $\bar{\mathbb{S}}$, (6) follows. Since every PVM is a POVM, $\bar{\mathbb{S}}$ is non empty for $k \geq N$.

Given the PVM P, by Gleason's theorem (Baras, 1979) $\exists \bar{\rho}_h \in \mathcal{T}_s^+(\mathbb{C}^{N\times k})$ such that $Tr[\bar{\rho}_h P(i)] = p_i^h$. Suppose there exists $\hat{\rho}_h$ such that $\bar{\rho}_h = V\hat{\rho}_h U$, then

$$p_i^h = Tr[\bar{\rho}_h P(i)] = Tr[V\hat{\rho}_h UP(i)] = Tr[\hat{\rho}_h M(i)]$$

Hence theorem 2 gives a possible approach to solve $P6$. Note that $\bar{\rho}_h = V\hat{\rho}_h U \Rightarrow \hat{\rho}_h = U\bar{\rho}_h V$, but $\hat{\rho}_h = U\bar{\rho}_h V \Rightarrow V\hat{\rho}_h U = VU\bar{\rho}_h VU$. By the given construction of V and U, $VU\bar{\rho}_h VU \neq \bar{\rho}_h$,. Hence $\hat{\rho}_h = U\bar{\rho}_h V$ is not a possible solution.

Consider the scenario described in the beginning of section 2.3. We describe a simple example of that scenario. Let

$\eta_1 = 3$ and $\eta_2 = 2$. When Y_2 is collected after Y_1, the distribution of the observations under hypothesis 0 and 1 is tabulated in the second and third columns of table 1 respectively. When Y_1 is collected after Y_2, the distribution of the observations under hypothesis 0 and 1 is tabulated in the fifth and sixth columns of table 1 respectively. The prior distribution of the hypothesis is set to $(\zeta_0 = 0.4, \zeta_1 = 0.6)$. The minimum probability of error when Y_2 is measured after Y_1 is 0.35. The minimum probability of error when Y_1 is measured after Y_2 is 0.266. Hence in this example the optimal strategy is first measure Y_2 and then measure Y_1.

$[Y_1, Y_2]$	$h = 0$	$h = 1$	$[Y_2, Y_1]$	$h = 0$	$h = 1$
1, 1	0.1	0.15	1, 1	0.25	0.15
1, 2	0.2	0.3	2, 1	0.05	0.30
2, 1	0.2	0.15	1, 2	0.25	0.13
2, 2	0.15	0.25	2, 2	0.1	0.27
3, 1	0.25	0.1	1, 3	0.05	0.12
3, 2	0.1	0.05	2, 3	0.3	0.03

Table 1. Distribution of observations under either hypothesis

4. CONCLUSION

In this paper we considered models for human decision making using classical (Kolmogorov) and non-classical (von Neumann, quantum like) probability models. We consider a simple decision making problem, the binary hypothesis testing problem with finite observation space. Using a particular state and PVM representation for the measurements, we formulated the detection problem with the objective of minimizing the probability of error. The solution to the detection problem was pure strategies and the expected cost with optimal strategies was the same as the minimum probability of error that could be achieved using classical probability models. Further, we proved that the minimum probability of error that can be achieved does not change with different state and PVM representations which achieve the distribution of the observations. In another approach, we represented the measurements using POVMs. We proved that the optimal strategies are pure strategies. Through an example, we discussed the effect of order in measurements.

As future work we are interested in considering other metrics which account for risk involved with a decision and low probability high impact events. Sequential hypothesis testing problems can be considered in VNPM framework. The trade off between the time to make decision, i.e., the number of observations collected to make a decision and the risk associated with the decision can be studied.

REFERENCES

Baras, J. and Harger, R. (1977). Quantum mechanical linear filtering of vector signal processes. *IEEE Transactions on Information Theory*, 23(6), 683–693.

Baras, J., Harger, R., and Park, Y. (1976). Quantum-mechanical linear filtering of random signal sequences. *IEEE Transactions on Information Theory*, 22(1), 59–64.

Baras, J.S. (1987). Distributed asynchonous detection: General models. In *Decision and Control, 1987. 26th IEEE Conference on*, volume 26, 1832–1835. IEEE.

Baras, J.S. (1988). An optimization problem from linear filtering with quantum measurements. *Applied Mathematics and Optimization*, 18(1), 191–214.

Baras, J. (1979). Noncommutative probability models in quantum communication and multi-agent stochastic control. *Richerche di Automatica*, 10(2), 21–265.

Borie, D. (2013). Expected utility theory with non-commutative probability theory. *Journal of Economic Interaction and Coordination*, 8(2), 295–315.

Buchanan, L. and O Connell, A. (2006). A brief history of decision making. *Harvard business review*, 84(1), 32.

Busch, P., Grabowski, M., and Lahti, P.J. (1997). *Operational quantum physics*, volume 31. Springer Science & Business Media.

Busemeyer, J.R. and Bruza, P.D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.

Busemeyer, J.R., Wang, Z., Townsend, J.T., and Eidels, A. (2015). *The Oxford handbook of computational and mathematical psychology*. Oxford University Press.

Davies, E.B. (1976). Quantum theory of open systems.

Frisch, D. and Clemen, R.T. (1994). Beyond expected utility: rethinking behavioral decision research. *Psychological bulletin*, 116(1), 46.

Hayashi, M. (2006). *Quantum information*. Springer.

Helstrom, C.W. (1969). Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2), 231–252.

Holevo, A.S. (2003). *Statistical structure of quantum theory*, volume 67. Springer Science & Business Media.

Khrennikov, A., Basieva, I., Dzhafarov, E.N., and Busemeyer, J.R. (2014). Quantum models for psychological measurements: an unsolved problem. *PloS one*, 9(10), e110909.

Neumann, J. (1955). *Mathematical foundations of quantum mechanics*. Princeton university press.

Trueblood, J.S. and Busemeyer, J.R. (2012). A quantum probability model of causal reasoning. *Frontiers in Psychology*, 3, 138.

Trueblood, J.S., Yearsley, J.M., and Pothos, E.M. (2017). A quantum probability framework for human probabilistic inference. *Journal of Experimental Psychology: General*, 146(9), 1307.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157), 1124–1131.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.

Von Neumann, J. and Morgenstern, O. (1947). Theory of games and economic behavior, 2nd rev.

Wang, Z., Solloway, T., Shiffrin, R.M., and Busemeyer, J.R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences*, 111(26), 9431–9436.