Statistics and Applications {ISSN 2454-7395 (online)}
Volume 19, No. 1, 2021 (New Series), pp 107–114

A Decision Theory in Non-commutative Domain

Kalyan B. Sinha
JN Centre for Advanced Scientific Research
Indian Institute of Science and
Indian Statistical Institute,
Bangalore, India

Received: 01 October 2020; Revised: 14 October 2020; Accepted: 01 November 2020

Abstract and Prologue

In the later part of my professional life in the Indian Statistical Institute (I.S.I.), when I left Delhi to take up the position of the Director of I.S.I. in Kolkata, Aloke was my pillar of support, my person-to-go-to in any crisis; his was the shoulder to cry on. Those were, in many ways, difficult times for me and often I reflect and wonder how those times would have been without Aloke. In fact, from the mid-nineties, Aloke, in his extremely pleasant, shy and humble way, slowly but surely entered into a very close friendship in my life, which I will cherish forever. My frequent travels to Delhi, well after my retirement from I.S.I., would bring me to I.S.I., Delhi and to me that meant spending time with Aloke, long discussions inevitably ending with a very pleasant lunch with him in the chinese restaurant, opposite the gate of I.S.I. All these will remain only memories now and my next visit to Delhi (delayed by Covid-19) will be empty, "Aloke-heen" (in Bengali), and will make me miss him all the more. The following article, a brief introduction (by a non-expert) to the decision theory in a non-commutative (quantum-) background, is my humble tribute to Aloke and to his friendship for me.

Key words: Decision theory; Quantum theory; Bayes decision rule.

AMS Subject Classifications: 62K05

1. Introduction

The statistical decision-theory or the idea of founding Statistics on a theory of decisions is due to Abraham Wald, enunciated in its originality, in his famous book, "Statistical Decision Functions" (Wald (1950), for a more recent account see the book of Ferguson (1967). There have been attempts, mainly by Holevo (see for example the books of Holevo (2011) and Hayashi (2017)), to recast these ideas in the context of non-commutative probabilistic background. As is well-known (see the first half of the book of Parthasarathy (1992) for an elegant account), the mathematical Quantum Theory represents a model of a non-Kolmogoroffian (or non-commutative) probability theory and hence there should be good reason to explore the possibility of studying an extension of the (classical) decision-theory to this domain. To give a brief account of this is the aim here.

Corresponding Author: Kalyan B. Sinha
Email: kbs@jncasr.ac.in

2. The Mathematical Description of The (Classical) Decision-Theory

As Ferguson (1967) observes in his book, the theory of games, as introduced by von Neumann in the 1940's, has a great deal of similarity with many aspects of decision theory. Both of these two theories start with three basic objects:
(i) a non-empty set of parameter, Θ, parametrizing the possible states of the system;
(ii) a non-empty set, Ω, of decisions (or actions) available to the statistician;
(iii) a function $L : \Theta \times \Omega \to \mathbb{R}$, called the **loss function** (the negative values of L needs to be interpreted as gain).

This triplet (Θ, Ω, L) defines a statistical decision problem or a game with the following interpretation. The nature (or providence!) chooses a point θ in Θ and the statistician, with no knowledge of the choice nature has made, makes a decision (or chooses an action) ω in Ω. As a consequence of these decisions, the statistician loses an amount $L(\theta, \omega)$. While in game-theoretic context, the players are trying simultaneously to minimize their losses, since the nature chooses the state without any such bias (hopefully!), this presents a dilemma for the decision-statistician and she tries to resolve this dilemma by gathering more information on the state by "sampling or by performing many experiments".

Thus for the decision-statistician, there is also a sample space $\mathcal{X}$ (here taken to be a Borel subset of $\mathbb{R}^d$, the d-dimensional Euclidean space) with a family of probability measures $\{\mu_\theta\}_{\theta \in \Theta}$ on $\mathcal{F}(\mathcal{X})$, the Borel σ-algebra of $\mathcal{X}$. The statistical decision problem, given by the triple (Θ, Ω, L) along with the sample space $\mathcal{X}$ of experiments, next chooses a (behavourial) decision map $D : \mathcal{X} \times \mathcal{F}(\Omega) \to \mathbb{R}_+$ such that $D(x, \cdot)$ is a probability measure on the Borel σ-algebra $\mathcal{F}(\Omega)$. Next one writes down the **risk function** $R : \Theta \times \{D\} \to \mathbb{R}$ by

$$R(\theta, D) = \int_\Omega L(\theta, \omega) \int_{\mathcal{X}} \mu_\theta(dx) D(x, d\omega). \tag{1}$$

An instructive way to rewrite (1) is to define the measure $\mu_\theta \circ D : \mathcal{F}(\Omega) \mapsto \mathbb{R}_+$ for every $\theta \in \Theta$ and $\Delta \in \mathcal{F}(\Omega)$ by

$$(\mu_\theta \circ D)(\Delta) = \int_{\mathcal{X}} \mu_\theta(dx) D(x, \Delta) \tag{2}$$

and replacing (1) by

$$R(\theta, D) = \int_\Omega L(\theta, \omega)(\mu_\theta \cdot D)(d\omega), \tag{3}$$

whenever the integral exists. Here we have noted that if $\mathcal{X} \ni x \mapsto D(x, \Delta)$ is measurable, then $\forall\ \theta$, $\mu_\theta \circ D$ is a probability measure on Ω and one can give a meaning to the integral in (3). The risk function R represents the average loss to the statistician when the nature has chosen the state parametrized by θ and the decision made is represented by the decision map D.

At this stage, one is still left with the problem of the "choice of parametrization" $\theta \in \Theta$ of the state and of the several avenues adopted by a statistician, we shall restrict our discussions here to the use of the "Bayes Principle". This involves putting a structure of a measure space on Θ and assigning a "prior probability measure"π on the σ-algebra $\mathcal{F}(\Theta)$. This leads naturally to the definition of the **Bayes risk of a (behavourial-)decision rule** D **with respect to the prior** π as

$$\begin{aligned}\mathcal{R}(\pi, D) &= \int_{\Theta} \pi(d\theta) R(\theta, D)\\ &= \int_{\Theta} \pi(d\theta) \int_{\Omega} L(\theta, \omega)(\mu_\theta \circ D)(d\omega). \end{aligned} \tag{4}$$

With regard to the definition (4), there are a few technical issues, *e.g.* the sense of measurability of the map $\theta \mapsto \mu_\theta \circ D$ etc., but these can be easily treated; for example in the above mentioned case one can have the assumption that $\theta \mapsto \mu_\theta(\cdot)$ is measurable and refer to [Dunford and Schwarz (1988), pages 156-162]. It is also worth mentioning that often authors (*e.g.* in Wald (1950)) consider the parameter space Θ to be finite or countably infinite. Also note, since all the 3 set-functions are non-negative, one can define a conditional probability measure of the random variable $\widehat{\theta}$ on Θ, given the random variable X on $\mathcal{X}$ (called the **posterior** probability measure of $\widehat{\theta}$ given the observation of X) on the product σ-algebra $\mathcal{F}(\Theta) \times \mathcal{F}(\mathcal{X})$ by

$$(\pi \cdot \mu)(\delta \times \Delta) = \int_{\delta} \pi(d\theta) \mu_\theta(\Delta) \tag{5}$$

for $\delta \in \mathcal{F}(\Theta), \Delta \in \mathcal{F}(\mathcal{X})$. In fact, in Ferguson (1967) the possibility of these two definitions (4) and (5) are pre-conditions for speaking about the "Bayes decision principle". This definition (5) sets up a linear ordering on the set $D(\cdot, \cdot)$ of decision functions and a Bayes decision rule is one that has the smallest Bayes risk, $\mathcal{R}$.

A decision function D_0 is said to be **Bayes with respect to the prior measure** π if

$$\mathcal{R}(\pi, D_0) = \inf_{D} \mathcal{R}(\pi, D). \tag{6}$$

It may happen that even if the right hand side of (6) exists, that value may **not** be attained for any D_0 and in such a case, one has to be satisfied with a decision D_0; which is "close" to the infimum. Let $\epsilon > 0$. A decision function $D_0 = D_0(\epsilon)$ is said to be ϵ**-Bayes** if

$$\mathcal{R}(\pi, D_0) \leq \inf_{D} \mathcal{R}(\pi, D) + \epsilon. \tag{7}$$

There are many other questions that arise naturally in the context of the above discussions; however, we shall take a break with the (classical) decision-theory and the rest of this article will be devoted to an attempt to "transport" the theory to the non-commutative (quantum) domain.

The definition (3) sets up a linear ordering (inherited from that of the real line) and the rule that is most preferred by that ordering is called the **minimax decision rule**: a decision map $D_0 \in \mathcal{D} \equiv$ the set $\{D : \mathcal{X} \times \mathcal{F}(\Omega) \to \mathbb{R}_+ \mid D(x, \cdot)$ is a probability measure with variation norm uniformly bounded w.r.t. $x \in \mathcal{X}\}$ is said to be minimax if

$$\sup_{\theta\in\Theta} R(\theta, D_0) = \inf_{D\in\mathcal{D}} \sup_{\theta\in\Theta} R(\theta, D). \tag{8}$$

If one assumes that (i) Θ and Ω are topological spaces such that Θ is compact, and $L : \Theta \times \Omega \to \mathbb{R}_+$ is continuous, (ii) $\Theta \ni \theta \mapsto \mu_\theta(\cdot)$ is continuous in w^*-topology of probability measures, then it can be seen that $\Theta \times \mathcal{D} \ni (\theta, D) \mapsto R(\theta, D)$ is continuous w.r.t the natural w^*-topology of $\mathcal{D}$, uniformly in θ. Therefore $\sup_\theta R(\theta, D)$ exists and $D \mapsto \sup_\theta R(\theta, D)$ is continuous w.r.t the w^*-topology of $\mathcal{D}$ in which $\mathcal{D}$ is compact. Thus the infimum exists and is attained since $\mathcal{D}$ is compact, *i.e.*, there exists a decision map D_0 with the property that

$$\inf_{D\in\mathcal{D}} \sup_{\theta\in\Theta} R(\theta, D) = \sup_{\theta\in\Theta} R(\theta, D_0).$$

A very similar proof for the partial quantum statistical decision rules can be constructed with $(\mu_\theta \circ D)(\cdot)$ replaced by $Tr_S(\rho_\theta D(\cdot))$ and very similar results can be obtained with same set of assumptions, as explained below.

3. Quantum Theory of Bayes' Decision-rules

If one thinks of the Quantum Theory as one possible model for non-Kolmogoroffian probability (see Partasarthy (1992) for an elaboration of this point of view), then the pair (sample space $\mathcal{X}$, real-valued random variable X) goes over to the relevant pair (Hilbert space h_S, a self-adjoint operator $\widehat{X}$ on it). Furthermore, the probability measure on $\mathcal{F}(X)$, associated with the random variable X is replaced by a density matrix ρ, a positive trace-class operator $(\mathcal{B}_{1+}(h_S))$ of trace 1, on h_S. In the present context of theory of decisions, there are two distinct possibilities:
(i) following Holevo's work (see Partasarthy (1992) and Holevo (1974)), one may have a kind of partial quantum (or non-commutative) statistical decision theory in which the sample space metamorphoses into its corresponding quantum structure, leaving the parameter-set Θ, a classical measure space with a prior probability measure π on it or (ii) a further or fully quantum statistical decision theory, in which the Bayesian part also undergoes a quantum metamorphosis. What turns out to be a remarkable coincidence (at least to the present author) that this second route has all the aspects of "quantum entanglement" (see. *e.g.* Petz (2008) and Parthasarthy (2013)) built in the mathematical structure.

For implementing the route (i), we first note that the sample space $\mathcal{X}$ is replaced by a (separable) Hilbert space h_S, the corresponding real-valued random variable X by a (possibly unbounded) self-adjoint operator $\widehat{X}$ in h_S and the family of probability measures $\{\mu_\theta(\cdot)\}_{\theta\in\Theta}$ by a family of density matrices $\rho \equiv \{\rho_\theta\}_{\theta\in\Theta} \in \mathcal{B}_{1+}(h_S)$ with $Tr_S(\rho_\theta) = 1$ for every θ, where Tr_S stands for the trace taken in the Hilbert space h_S. Furthermore, the triple (Θ, Ω, L) are given as before with Θ and Ω as two measure spaces and $L : \Theta \times \Omega \to \mathbb{R}_+$ measurable loss function. The most important change that takes place here is the replacement of the (behavourial) decision-function $D(x, \Delta)$ (for $x \in \mathcal{X}, \Delta \in \mathcal{F}(\Omega)$) by a map $D : \mathcal{F}(\Omega) \to \mathcal{B}_+(h_S)$, the set of non-negative bounded operators on h_S such that it is countably additive: $\{\Delta_j\}_{j=1}^\infty$ of disjoint subsets in $\mathcal{F}(\Omega)$ such that $\Delta = \bigcup_{j=1} \Delta_j$ implies that

$D(\Delta) = \sum_{j=1}^{\infty} D(\Delta_j)$ (the infinite sum converging in strong operator topology), and $D(\Omega) = I \in \mathcal{B}(h_S)$. This kind of family is called a POVM (positive operator-valued measures) on Ω (see Holevo (2011) and Davies (1976) for some applications of POVM). We can now define the partial quantum risk function (p.q.r.f) as:

$$R(\theta, D) = \int L(\theta, \omega) Tr_S(\rho_\theta D(d\omega)). \tag{9}$$

The right hand side makes sense since the map $\rho \circ D : \Theta \times \mathcal{F}(\Omega) \mapsto \mathbb{R}_+$ given by

$$(\rho \circ D)(\theta, \Delta) = Tr_S(\rho_\theta D(\Delta)) = Tr_S(\rho_\theta^{1/2} D(\Delta) \rho_\theta^{1/2}), \tag{10}$$

is a non-negative countably additive set-function with $(\rho \circ D)(\theta, \Omega) = 1$ and hence defines a probability measure on Ω for every $\theta \in \Theta$. Thus (9) makes sense as a Lebesgue integral and (9) can be rewritten as

$$R(\theta, D) = \int L(\theta, \omega)(\rho \cdot D)(\theta, d\omega). \tag{11}$$

Finally, with the prior probability measure π on $\mathcal{F}(\Theta)$, one has as in (4), the partial quantum Bayes' risk (p.q.B.r) of a (behaviourial) decision rule D:

$$\begin{aligned} \mathcal{R}(\pi, D) &= \int_\Theta \pi(d\theta) R(\theta, D) \\ &= \int_\Theta \pi(d\theta) \int_\Omega L(\theta, \omega)(\rho \cdot D)(\theta, d\omega). \end{aligned} \tag{12}$$

As in the classical case, one can define a (partially quantum) conditional density matrix of the random variable $\widehat{\theta}$ on Θ, given the (quantum) observation of the operator $\widehat{X}$ in h_S (we shall call it as **posterior density matrix** of $\widehat{\theta}$ given $\widehat{X}$ in h_S):

$$(\pi \cdot \rho)(\delta) = \int_\delta \pi(d\theta) \rho_\theta, \ \forall \ \delta \in \mathcal{F}(\Theta), \tag{13}$$

where the integral on the right hand side is the strong Bochner integral in the Banach space $\mathcal{B}_1(h_S)$. It is easy to see that this $\mathcal{B}_{1+}(h_S)$-valued set function on $\mathcal{F}(\Theta)$ is countably additive and $Tr_S(\pi \cdot \rho)(\Theta) = 1$. In fact, the Bayes risk p.q.B.r can be rewritten in terms of the posterior density matrix $(\pi \circ \rho)(\cdot)$ as

$$\mathcal{R}(\pi, D) = \int_{\Theta \times \Omega} L(\theta, \omega) Tr_S((\pi \cdot \rho)(d\theta) D(d\omega)). \tag{14}$$

In analogy, given a prior π, the partially quantum Bayes decision rule is the D which gives the smallest p.q.B.r and a decision D_0 (in $\mathcal{B}_+(h_S)$-valued POVM's on $\mathcal{F}(\Omega)$) is said to be **Bayes with respect to prior** π if

$$\mathcal{R}(\pi, D_0) = \inf_{D \in h_S - povm(\Omega)} \mathcal{R}(\pi, D). \tag{15}$$

In the rest of this article, we consider fully quantum decision theory in which the sample space $\mathcal{X}$ as well as parameter space Θ metamorphoses into two (separable) Hilbert space h_S and h_B, respectively, and $\pi \cdot \mu(\cdot)$ or $\pi \cdot \rho$ are replaced by one density matrix Φ on $\tilde{h} = h_S \otimes h_B$. This structure, in conjunction with the following assumptions constitute the present new proposal.

A1. Ω is a compact Borel space and the loss operator $L : \Omega \to \mathcal{B}_+(h_B)$ is continuous w.r.t the ω^*-topology of $\mathcal{B}(h_B)$;

A2. $D : \mathcal{F}(\Omega) \mapsto \mathcal{B}_+(h_S)$ is a POVM, as mentioned earlier and as in Holevo's theory.

Then we lift these two operator-families to the Hilbert space $\tilde{h}$ by setting

$$\begin{aligned} \tilde{L}(\omega) &= I_S \otimes L(\omega) \text{ for } \omega \in \Omega \text{ and} \\ \tilde{D}(\Delta) &= D(\Delta) \otimes I_B \text{ for } \Delta \in \mathcal{F}(\Omega). \end{aligned} \tag{16}$$

Note that $\tilde{L}(\omega)$ commutes with $\tilde{D}(\Delta)$ in $\tilde{h}$ and we define the **fully quantum risk function**

$$\mathcal{R}(\Phi, D) = \int_\Omega Tr_{\tilde{h}}(\Phi \tilde{L}(\omega)\tilde{D}(d\omega)), \tag{17}$$

which is

$$= \int_\Omega Tr_{\tilde{h}}\{(\tilde{L}(\omega)^{1/2}\Phi\tilde{L}(\omega)^{1/2})\tilde{D}(d\omega)\},$$

showing that $\mathcal{R}(\Phi, D) \geq 0$, if it exists. The issue of the sense in which the integral in (17) can be defined is not a simple one and it is left unresolved in this article, to be dealt with later. However, it should be mentioned that Holevo (see *e.g.* Holevo (1974)) gave a theory to study such integrals. Here we shall restrict ourselves to the simpler case when the density matrix Φ on $\tilde{h}$ is a **finite** linear combination of tensors of density matrices on h_S and h_B:

$$\Phi = \sum_{j=1}^n \rho_j \otimes \pi_j; \;\; \rho_j \in \mathcal{B}_{1+}(h_S), \pi_j \in \mathcal{B}_{1+}(h_B). \tag{18}$$

In such a case,

$$\Phi\tilde{L}(\omega)\tilde{D}(d\omega) = \sum_{j=1}^n (\rho_j D(d\omega) \otimes (\pi_j L(\omega))$$

and thus we shall be looking at the "integral"

$$\int_\Omega Tr_B(\pi_j L(\omega)) \cdot Tr_S(\rho_j D(d\omega)), \tag{19}$$

which exists as a Lebesgue-type integral since the function $Tr_S(\pi_j L(\cdot))$ is bounded continuous on compact Ω and since the second factor in (19) is clearly a (non-negative) finite measure with total variation $= Tr_S(\rho_j)$. For the rest of the discussion, viz. the one on a kind of minimax theorem, we shall assume that the integral in (17) exists for all density matrices Φ on $h_S \otimes h_{\mathcal{B}}$.

As we have observed before, by virtue of the assumption A1, the map: density matrices on $\tilde{h} \ni \Phi \mapsto \mathcal{R}(\Phi, D) \in \mathbb{R}_+$ is continuous w.r.t. the w^*-topology on density matrices induced by $\mathcal{B}(\tilde{h})$ after applications of Mazur's theorem. Also note that Alaoglu's theorem implies that in the same topology, the set of density matrices is a (convex) compact set and therefore, there exists a density matrix Φ_0 such that

$$\sup_{\Phi} \mathcal{R}(\Phi, D) = \mathcal{R}(\Phi_0, D). \tag{20}$$

On the other hand, it is easy to see that

$$\begin{aligned} \sup_{\Phi} \inf_{D} \mathcal{R}(\Phi, D) &\leq \inf_{D} \sup_{\Phi} \mathcal{R}(\Phi, D) \\ &= \inf_{D} \mathcal{R}(\Phi_0, D) \\ &\leq \sup_{\Phi} \inf_{D} \mathcal{R}(\Phi, D) \end{aligned}$$

and therefore one has

$$\sup_{\Phi} \inf_{D} \mathcal{R}(\Phi, D) = \inf_{D} \sup_{\Phi} \mathcal{R}(\Phi, D). \tag{21}$$

The left hand side is called the lower value and the right hand side the upper value and equality of these two constitutes the **minimax** decision rule.

The procedure and results, indicated above can be strengthened more, in line with the classical case, if instead we ask the following:

Given $\sigma \in \mathcal{B}_{1+}(h_B)$, let $\mathcal{S}_\sigma = \{\Phi \in \mathcal{B}_{1+}(h_S \otimes h_B) \mid Tr_S \Phi = \sigma\}$.

Then does there exist a POVM D_0 such that $\sup\limits_{\Phi \in \mathcal{S}_\sigma} \inf\limits_{D} \mathcal{R}(\Phi, D) = \sup\limits_{\Phi \in \mathcal{S}_\sigma} \mathcal{R}(\Phi, D_0)$?

Acknowledgement

This endeavour is partly supported by the I.N.S.A. Senior Scientist programme of the Indian National Science Academy, for which the author is grateful.

References

Davies, E. B. (1976). *Quantum Theory of Open Systems.* Academic Press, New York.

Dunford, N. and Schwartz, J. T. (1988). *Linear Operators - Part I.* John Wiley, USA.

Ferguson, T. (1967). *Mathematical Statistics - A Decision Theoretic Approach.* Academic Press, New York.

Hayashi, M. (2017). *Quantum Information Theory- Mathematical Foundations*, Second Edition, Springer-Verlag, Berlin.

Holevo, A. S. (1974). The theory of statistical decisions on an operator algebra. *Dokaldy Akademii Nauk SSSR*, **218(1)**, 54–57.

Holevo, A. S. (2011). *Probabilistic and Statistical Aspects of Quantum Theory*, Second Edition. Edizioni Della Normale, Pisa.

Parthasarathy, K. R. (1992). *Introduction to Quantum Stochastic Calculus.* Birkhauser Verlag, Basel.

Parthasarathy, K. R. (2013). *Coding Theorems of Classical and Quantum Information Theory.* TRIM series, HBA, New Delhi.

Petz, D. (2008). *Quantum Information Theory and Quantum Statistics.* Springer Verlag, Berlin.

Wald, A. (1950). *Statistical Decision Functions.* Wiley, New York.