Introduction to Hilbert Space Multi-Dimensional Modeling

Jerome Busemeyer and Zheng Joyce Wang

Abstract This chapter provides a brief introduction to procedures for estimating Hilbert space multi-dimensional (HSM) models from data. These models, which are built from quantum probability theory, are used to provide a simple and coherent account of a collection of contingency tables. The collection of tables are obtained by measurement of different overlapping subsets of variables. HSM models provide a representation of the collection of the tables in a low dimensional vector space, even when no single joint probability distribution across the observed variables can reproduce the tables. The parameter estimates from HSM models provide simple and informative interpretation of the initial tendencies and the inter-relations among the variables.

Keywords Quantum probability · Non-commutativity · Data fusion

1 Introduction

This chapter provides an introduction to computational tools, based on what we call Hilbert space multi-dimensional theory, which can be used for representing data tables from multiple sources by a single coherent vector space and linear operations on the space. For more complete descriptions of this theory, see the original articles by the authors Busemeyer and Wang [8, 9], which include detailed worked out examples. Here we plan to outline the main steps of building a program and also point to computer programs available to process collections of data tables.

J. Busemeyer (✉)
Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA
e-mail: jbusemey@indiana.edu

Z. J. Wang
School of Communication, Center for Cognitive and Brain Sciences, The Ohio State University, Columbus, OH, USA

D. Aerts et al. (eds.), *Quantum-Like Models for Information Retrieval and Decision-Making*, STEAM-H: Science, Technology, Engineering, Agriculture, Mathematics & Health, https://doi.org/10.1007/978-3-030-25913-6_3

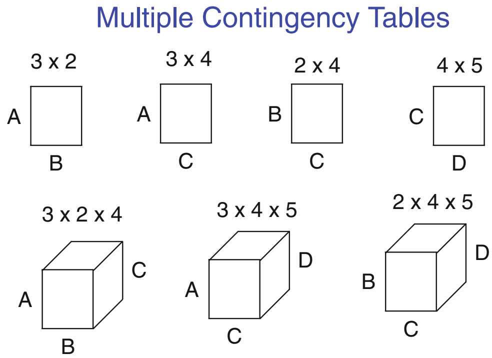

Fig. 1 Illustration of a collection of contingency data tables

Figure 1 illustrates the basic problem that we wish to address. Suppose large medical data sites provide information about co-occurrence of various kinds of symptoms, labeled A, B, C, and D in the figure. The symptoms can be manifest to different degrees. For example, symptom B is binary valued, symptom A has three levels, symptom C has four degrees, and symptom D has five rating values. Suppose different methods for querying the sites yield different contingency tables summarizing co-occurrence of pairs of variables and co-occurrence of triples of variables, which produce the tables shown in the figure. The cells of the contingency tables contain the frequency of a combination of symptoms. For example, the A by B by C table is a 3 by 2 by 4 table, and each cell contains the frequency that a particular combination of values was assigned to the variables A, B, C, using one query method.

The following problem arises from considering all these contingency tables. How does a data scientist integrate and synthesize these seven different tables into a compressed, coherent, and interpretable representation? This is a problem that often arises in relational database theory [1]. It is common to apply categorical data analysis [3] to a single table (e.g., a single A by B by C by D table). However, the problem is different here because there are a collection of seven tables of varying dimensions rather than a single four-way table. Alternatively, one could try Bayesian networks, which require assuming that all the tables are generated from a *single* latent four-way joint distribution [10]. Unfortunately, however, it may be the case that no four-way joint distribution exists that can reproduce all the observed tables! This occurs when the data tables violate consistency constraints, forced by marginalization, upon which Bayes nets rely to fit the tables [11].

Hilbert space multi-dimensional (HSM) models are based on quantum probability theory [13]. They provide a way to account for a collection of tables, such as illustrated in Fig. 1, even when no four-way joint distribution exists. HSM models provide an estimate of the target population's initial tendencies in the form of a state vector, and HSM models represent the inter-relationships between the different variables (symptoms in this example) using "rotations" of the basis of the vector space.

This chapter is organized as follows: First, we summarize the basic principles of quantum probability theory, then we summarize the steps required to build an HSM model, and finally we refer to programs available on the web for applying an HSM model to real data.

2 Basics of Quantum Probability Theory

The idea of applying quantum probability to the field of judgment began from several directions [2, 4, 5, 14]. The first to apply these ideas to the field of information retrieval was by van Rijsbergen [18]. For more recent developments concerning the application of quantum theory to information retrieval, see [16]. van Rijsbergen argues that quantum theory provides a sufficiently general yet rigorous formulation for integration of three different approaches to information retrieval—logical, vector space, and probabilistic. Another important reason for considering quantum theory is that human judgments (e.g., judging presence of symptoms by doctors) have been found to violate rules of Kolmogorov probability, and quantum probability provides a formulation for explaining these puzzling findings (see, e.g., [7]).

HSM models are based on quantum probability theory and so we need to briefly review some of the basic principles used from this theory.[1]

In quantum theory, a variable (e.g., variable A) is called an observable, which corresponds to the Kolmogorov concept of a random variable. The pair of a measurement of a variable and an outcome generated by measuring a variable is an event (e.g., measurement of variable A produces the outcome 3, so that we observe $A = 3$).

Kolmogorov theory represents events as subsets of a sample space, Ω. Quantum theory represents events as subspaces of a Hilbert space H.[2] Each subspace, such as A, corresponds to an orthogonal projector, denoted P_A for subspace A. An orthogonal projector is used to project vectors into the subspace it represents.

In Kolmogorov theory, the conjunction "A and B" of two events, A and B, is represented by the intersection of the two subsets representing the events (e.g., $(A = 3) \cap (B = 1)$). In quantum theory, a sequence of events, such as A and then B, denoted AB, is represented by the sequence of projectors $P_B P_A$. If the projectors commute, $P_A P_B = P_B P_A$, then the product of the two projectors is a projector corresponding to the subspace $A \cap B$, that is, $P_B P_A = P(A \cap B)$; and the events A and B are said to be *compatible*. However, if the two projectors do not commute, $P_B P_A \neq P_A P_B$, then neither their product is a projector, and the events

[1] See [7, 15, 16, 18] for tutorials for data and information scientists.

[2] Technically, a Hilbert space is a complex valued inner product vector space that is complete. Our vectors spaces are finite, and so they are always complete.

are *incompatible*. The concept of incompatibility is a new contribution of quantum theory, which is not present in Kolmogorov theory.

Kolmogorov theory defines a state as a probability measure p that maps events to probabilities. Quantum theory uses a unit length state vector, here denoted ψ, to assign probabilities to events.[3] Probabilities are then computed from the quantum algorithm

$$p(A) = \|P_A \psi\|^2 . \tag{1}$$

Both Kolmogorov and quantum probabilities satisfy the properties for an additive measure. In the Kolmogorov case, $p(A) \geq 0$, $p(\Omega) = 1$, and if $(A \cap B) = 0$, then $p(A \cup B) = p(A) + p(B)$.[4] In the quantum case, $p(A) \geq 0$, $p(H) = 1$, and if $P_A P_B = 0$, then $p(A \vee B) = p(A) + p(B)$.[5] In fact, Eq. (1) is the unique way to assign probabilities to subspaces that form an additive measure for dimensions greater than 2 [12].

Kolmogorov theory defines a conditional probability function as follows:

$$p(B|A) = \frac{p(A \cap B)}{p(A)},$$

so that the joint probability equals $p(A \cap B) = p(A)p(B|A) = p(B)p(A|B) = p(B \cap A)$, and order does not matter. In quantum theory, the corresponding definition of a conditional probability is

$$p(B|A) = \frac{\|P_B P_A \psi\|^2}{p(A)},$$

and so the probability of the sequence AB equals $p(AB) = p(A) \cdot p(B|A) = \|P_B P_A \psi\|^2$. In the quantum case, this definition of conditional probability incorporates the property of incompatibility: if the projectors do not commute, so that $P_A P_B \neq P_B P_A$, then $p(AB) \neq p(BA)$, and order of measurement matters. Extensions to sequences with more than two events follow the same principles for both classical and quantum theories. For example, the quantum probability of the sequence $(AB)C$ equals $\|P_C\,(P_B P_A)\,\psi\|^2$.

[3] A more general approach uses what is called a density operator rather than a pure state vector, but to keep ideas simple, we use the latter.

[4] $\cup$ is the union of subsets A, B.

[5] $\vee$ is the span of subspaces A, B.

3 Steps to Build an HSM Model

An HSM model is constructed from the following six steps:

1. Determine the compatibility and incompatibility relations among the variables.
2. Determine the dimension of the Hilbert space based on assumed compatibility relations.
3. Define the initial state given the dimension of the Hilbert space.
4. Define the projectors for the variables using unitary transformations to change the basis.
5. Compute the choice probabilities given the initial state and the projectors.
6. Estimate model parameters, compare fit of models.

3.1 How to Determine the Compatibility Relations

There are two ways to investigate and determine compatibility between a pair of variables. The direct way is to empirically determine whether or not the joint frequencies change depending on order of presentation. If there are order effects, then that is evidence for incompatibility. An indirect way is to compare fits of models that assume different compatibility relations. This indirect methods might be needed if no empirical tests of order effects are available.

3.2 How to Determine the Dimension

The basic idea of HSM modeling is to start with the minimum dimension required, and then add dimensions only if needed to obtain a satisfactory fit to the data. Of course this model comparison and model selection process needs to provide a reasonable balance between accuracy and parsimony. For example, when fitting the models using maximum likelihood estimation, model comparison indices such as Bayesian information criterion or Akaike information criterion can be used.

The minimum dimension is determined by the maximum number of combinations of values produced by the maximum number of compatible variables. For example, in Fig 1, suppose variables B and C are compatible with each other, and variables A and D are compatible with each other, but the pair B,C is incompatible with the pair A,D. In this case, there are at most two compatible variables. The B,C pair produces $2 \cdot 4 = 8$ combinations of values, but the A,D pair produces $3 \cdot 5 = 15$ combinations. The minimum dimension needs to include all 15 combinations produced by the A,D pair. Therefore, the minimum dimension is 15 in this example.

3.3 Define the Initial State

The compatible variables can be chosen to form the basis used to define the coordinates of the initial state ψ. In this example, the space is 15 dimensional, and the compatible pair, A,D can be chosen to define the initial basis for the unit length 15×1 column matrix ψ. Each coordinate ψ_{ij} represents the amplitude corresponding to the pair of values $(A = i, D = j)$, $i = 1, 2, 3$; $j = 1, 2, \ldots, 5$, for representing the initial state. The squared magnitude of a coordinate equals the probability of the combination, $p(A = i, D = j) = \left|\psi_{ij}\right|^2$. In general, the coordinates can be complex, but in practice they are usually estimated as real values.

3.4 Define the Projectors

The orthogonal projector for an event that is defined in the initial basis is simply an indicator matrix that picks out the coordinates that correspond to the event. For example, using the previous example, the projector for the event $(A = i, D = j)$ is simply $P_{A=i,D=j} = diag\left[0 \cdots 1 \cdots 0\right]$, where the one is located in the row corresponding to (i, j), which is a one-dimensional projector. The projector for the event $(A = i)$ equals $P_{A=i} = \sum_j P_{A=i,D=j}$ and the projector for the event $(D = j)$ equals $P_{D=j} = \sum_i P_{A=i,D=j}$, and these are multi-dimensional projectors.

The projectors for the incompatible events require changing the basis from the original basis to the new basis for the incompatible variables. For example, suppose we wish to define the events for the variables B,C. If we originally defined the initial state ψ in the B,C basis from the start, then these projectors would simply be defined by indicator matrices as well. Recall that the dimension of the space is 15, and there are only 8 combination of values for B, C. Therefore one or more of the combinations for B, C need to be defined by a multi-dimensional projector, M_{kl}, which is simply an indicator matrix, such as $M_{kl} = diag\left[1\ 0 \ldots 1\ 0\right]$ that picks out two or more coordinates for the event $(B = k, C = l)$. The collection of indicator matrices, $\{M_{kl}, k = 1, 2; l = 1, 2, 3, 4\}$, forms a complete orthonormal set of projectors. The projector for the event $(B = k)$, in the B, C basis, is simply $p(B = k) = \sum_l M_{B=k,C=l}$, and the projector for $(C = l)$ in the B, C basis is $p(C = l) = \sum_k M_{B=k,C=l}$.

We did not, however, define the initial state ψ in terms of the B,C basis. Instead we defined the initial state ψ in terms of the A,D basis. Therefore we need to "rotate" the basis from the A, D basis to the B, C basis to form the B,C projectors as follows: $P_{B=k,C=l} = U \cdot M_{jk} \cdot U^\dagger$, where U is a unitary matrix (an orthonormal matrix). Now the projector for the event $(B = k)$, in the A, D basis, is $p(B = k) = \sum_l P_{B=k,C=l}$, and the projector for $(C = l)$ in the A, D basis is $p(C = l) = \sum_k P_{B=k,C=l}$.

Any unitary matrix can be constructed from a Hermitian matrix $H = H^\dagger$ by the matrix exponential $U = exp(-i \cdot H)$. Therefore, the most challenging problem is to construct a Hermitian matrix that captures the change in bases. This is facilitated by using substantive theory from the domain under investigation. We describe this step in more detail in the original articles.

3.5 Compute the Choice Probabilities

Once the projectors have been defined, it is straightforward to compute the probabilities for any contingency table using the quantum algorithm described earlier. For example, the probabilities for the AB table are obtained from the equation $p(A = i, B = j) = \left\| P_{B=j} P_{A=i} \psi \right\|^2$, and the probabilities for the A, B, D table are obtained from the equation $p(A = i, B = k, D = j) = \left\| P_{D=j} P_{B=k} P_{A=i} \cdot \psi \right\|^2$.

3.6 Estimate Model Parameters, Compare and Test Models

Once the model has been defined, the parameters of the initial state ψ and the parameters in the Hamiltonian matrix H can be estimated from the frequencies contained within contingency table data. This can be accomplished by using maximum likelihood estimation procedures. Suppose the dimension equals n ($n = 15$ in our example). If we use a real valued initial state, then initial state has $n - 1$ parameters (because the state is restricted to unit length). If the Hamiltonian is restricted to real values, then the Hamiltonian has $(n \cdot (n + 1)/2) - 1$ parameters (one diagonal entry is arbitrary). However, often it is possible to use a lower number of parameters for the Hamiltonian. Model comparison methods, such as Bayesian information criterion or Akaike information criterion, can be used to compare models for accuracy adjusted for parsimony (defined by number of parameters).

HSM models can also be empirically tested using a generalization criterion. After estimating the projectors from two-way tables shown in Fig. 1, the model can be used to make a priori predictions for table A by D or for a three-way table such as A by B by D. This provides strong empirical tests of the model predictions.

The model also provides interpretable parameters to help understand the complex array of contingency tables. The estimate of the initials state ψ provides the initial tendencies to respond to questions. In the previous example, ψ represents the probabilities to respond to the A, D questions. The rotation, $U^\dagger \psi$ gives the initial tendencies to respond to the B, C questions. The squared magnitude of an entry in the unitary matrix, $\left|u_{jk}\right|^2$, represents the squared correlation between a basis vector representing an event in one basis (e.g., an event in the A, D basis) and a basis vector representing an event in another basis (e.g., the B, C basis). These squared correlations describe the inter-relations between the variables, independent of the initial state.

4 Computer Programs

We have started developing computer programs for fitting HSM models to different collections of contingency tables. These programs are currently located at the following site http://mypage.iu.edu/~jbusemey/quantum/HilbertSpaceModelPrograms.htm.

The site contains a link to some commonly used programs required for all of the models. It also contains programs designed to fit (a) collections of one and two-way tables made from binary variables, such as those that appear in [9], (b) one and two-way tables for variables with 2, 3, 4 values, such as those that appear in [8], (c) a model for order effects between a pair of variables with a relatively large (e.g., nine or greater) levels of rating scale values, such as those that appear in [19].

5 Concluding Comments

HSM models provide a simple and low dimensional method for representing multiple contingency tables formed from measurement of subsets of variables. This simple representation in low dimensional spaces is achieved by using "rotation" of the basis vectors to generate new incompatible variables. Bayesian network models can also be applied to collections of tables; however, these types of models assume the existence of a complete joint distribution of the observed variables, and it is often the case that no complete joint distribution can reproduce the tables because of violations of constraints imposed by marginalization. HSM models can be applied to collections of tables even when no complete joint distribution exists to reproduce the collection. Of course, HSM models do not provide the only way, and there are other probabilistic models that could be considered such as the use of probabilistic data base programming methods [6]. However, HSM models have been shown to provide successful accounts of actual empirical data [8, 9, 17, 19], as well as the possibility for providing new a priori predictions for new data, which is not the case for probabilistic database programming methods.

Acknowledgements This research was based upon the work supported by NSF SES-1560501 and NSF SES-1560554.

References

1. Abramsky, S. (2013). Relational databases and Bells theorem. In V. Tannen, L. Wong, L. Libkin, W. Fan, T. Wang-Chiew, & M. Fourman (Eds.), *In search of elegance in the theory and practice of computation* (pp. 13–35). Berlin: Springer.
2. Aerts, D., & Aerts, S. (1995). Applications of quantum statistics in psychological studies of decision processes. *Foundations of Science, 1*(1), 85–97.

3. Agresti, A., & Katera, M. (2011). *Categorical data analysis*. Berlin: Springer.
4. Atmanspacher, H., Römer, H., & Walach, H. (2002). Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics, 32*(3), 379–406.
5. Bordley, R. F. (1998). Quantum mechanical and human violations of compound probability principles: Toward a generalized Heisenberg uncertainty principle. *Operations Research, 46*(6), 923–926.
6. Bruza, P. D., Kitto, K., Ramm, B. J., & Sitbon, L. (2015). A probabilistic framework for analysing the compositionality of conceptual combinations. *Journal of Mathematical Psychology, 67*, 26–38.
7. Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge: Cambridge University Press.
8. Busemeyer, J. R., & Wang, Z. (2018). Data fusion using Hilbert space multi-dimensional models. *Theoretical Computer Science, 752*, 41–55.
9. Busemeyer, J. R. & Wang, Z. (2018). Hilbert space multidimensional theory. *Psychological Review, 125*(4), 572–591.
10. Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks*. New York, NY: Cambridge University Press.
11. Dzhafarov, E., & Kujala, J. V. (2012). Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychology, 56*, 54–63.
12. Gleason, A. M. (1957). Measures on the closed subspaces of a Hilbert space. *Journal of Mathematical Mechanics, 6*, 885–893.
13. Gudder, S. P. (1988). *Quantum probability*. Boston, MA: Academic Press.
14. Khrennikov, A. (1999). Classical and quantum mechanics on information spaces with applications to cognitive, psychological, social, and anomalous phenomena. *Foundations of Physics, 29*(7), 1065–1098.
15. Khrennikov, A. Y. (2010). *Ubiquitous quantum structure: From psychology to finance*. Berlin: Springer.
16. Melucci, M. (2015). *Introduction to information retrieval and quantum mechanics*. Berlin: Springer.
17. Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological Review, 120*(3), 679–696
18. van Rijsbergen, C. J. (2004). *The geometry of information retrieval*. Cambridge: Cambridge University Press.
19. Wang, Z., & Busemeyer, J. R. (2016). Comparing quantum versus Markov random walk models of judgments measured by rating scales. *Philosophical Transactions of the Royal Society, A, 374*, 20150098.