# A quantum framework for likelihood ratios

Article (Accepted Version)

International Journal of Quantum Information


# A QUANTUM FRAMEWORK FOR LIKELIHOOD RATIOS

RACHAEL L. BOND

*School of Psychology, University of Sussex, Falmer, East Sussex, BN1 9QH, United Kingdom*
*hello@rachaelbond.com*

YANG-HUI HE

*Department of Mathematics, City, University of London, EC1 0HB, United Kingdom, and*
*Merton College, University of Oxford, OX1 4JD, United Kingdom, and*
*School of Physics, NanKai University, Tianjin, 300071, P.R. China*
*hey@maths.ox.ac.uk*

THOMAS C. ORMEROD

*School of Psychology, University of Sussex, Falmer, East Sussex, BN1 9QH, United Kingdom*
*t.ormerod@sussex.ac.uk*

Received 6, March 2017
Revised 3, November 2017

The ability to calculate precise likelihood ratios is fundamental to science, from Quantum Information Theory through to Quantum State Estimation.However, there is no assumption-free statistical methodology to achieve this. For instance, in the absence of data relating to covariate overlap, the widely used Bayes theorem either defaults to the marginal probability driven "naive Bayes' classifier", or requires the use of compensatory expectation-maximization techniques. This article takes an information-theoretic approach in developing a new statistical formula for the calculation of likelihood ratios based on the principles of quantum entanglement, and demonstrates that Bayes' theorem is a special case of a more general quantum mechanical expression.

*Keywords*: Bayes' theorem; Probability; Statistics; Inference; Decision-making.

## Contents

## 1. Introduction

In recent years, Bayesian statistical research has often been epistemologically driven, guided by de Finetti's famous quote that "probability does not exist".[1] For example, the "quantum Bayesian" methodology of Caves, Fuchs, & Schack has applied de Finetti's ideas to Bayes' theorem for use in quantum mechanics.[2] In doing so, Caves et al. have argued that statistical systems are best interpreted by methods in which the Bayesian likelihood ratio is seen to be both external to the system and subjectively imposed on it by the observer.[3] However, the Caves et al. approach is problematic. At a human scale, for instance, an observer's belief as to the chances of a fair coin landing either "heads" or "tails" has no known effect. Indeed, for all practical purposes, the "heads:tails" likelihood ratio of 0.5:0.5 is only meaningful when considered as a property of the coin's own internal statistical system rather than as some ephemeral and arbitrary qualia.

Yet, to date, the axiomatic difficulties associated with Bayes' theorem, notably its reliance upon the use of marginal probabilities in the absence of structural statistical information (eg., estimates of covariate overlap), as well as the assumed conditional independence of data, have largely been approached from a "mend and make do" standpoint. For instance, the "maximum likelihood" approach of Dempster, Laird, & Rubin calculates iteratively derived measures of covariate overlap which not only lack a sense of natural authenticity, but also introduce fundamental assumptions into the statistical analysis.[4]

Instead, this paper adopts a different approach to the analysis of statistical systems. By using quantum mechanical mathematical spaces, it is demonstrated that the creation of isomorphic representations of classical data-sets as entangled systems allows for a natural, albeit non-trivial, calculation of likelihood ratios. It is expected that this technique will find applications within the fields of quantum state estimation, and quantum information theory.

## 2. The limits of Bayes' theorem

Bayes' theorem is used to calculate the conditional probability of a statement, or hypothesis, being true given that other information is also true. It is usually written as

$$P(H_i|D) = \frac{P(H_i)P(D|H_i)}{\sum_j P(H_j)P(D|H_j)} \ . \tag{2.1}$$

Here, $P(H_i|D)$ is the conditional probability of hypothesis $H_i$ being true given that the information $D$ is true; $P(D|H_i)$ is the conditional probability of $D$ being true if $H_i$ is true; and $\sum_j P(H_j)P(D|H_j)$ is the sum of the probabilities of all hypotheses

multiplied by the conditional probability of $D$ being true for each hypothesis.[5]

| | Particle $\alpha$ ($H_1$) | Particle $\beta$ ($H_2$) |
|---|---|---|
| Number of particles ($n$) | 10 | 10 |
| Proportion spin $\uparrow$ ($D$) | 0.8 | 0.7 |
| Proportion spin $\downarrow$ ($\bar{D}$) | 0.2 | 0.3 |

(2.2)

To exemplify using the contingency information in (2.2), if one wishes to calculate the nature of a randomly selected particle from a set of 20, given that it has spin $\uparrow$, then using Bayes' theorem it is trivial to calculate that particle $\alpha$ is the most likely type with a likelihood ratio of approximately 0.53:0.47,

$$\begin{aligned} P(H_1|D) &= \frac{0.5 \times 0.8}{(0.5 \times 0.8) + (0.5 \times 0.7)} = \frac{8}{15} && \approx 0.533 \ , \\ P(H_2|D) &= 1 - P(H_1|D) = \frac{7}{15} && \approx 0.467 \ , \end{aligned} \tag{2.3}$$

where $P(H_i) = 10/(10 + 10) = 0.5$ for both $i = 1, 2$.

However, difficulties arise in the use of Bayes' theorem for the calculation of likelihood ratios where there are multiple non-exclusive data sets. For instance, if the information in (2.2) is expanded to include data about particle charge (2.4) then the precise covariate overlap (i.e., $D_1 \cap D_2$) for each particle becomes an unknown.

| | Particle $\alpha$ ($H_1$) | Particle $\beta$ ($H_2$) |
|---|---|---|
| Number of particles ($n$) | 10 | 10 |
| Proportion spin $\uparrow$ ($D_1$) | 0.8 | 0.7 |
| Proportion charge + ($D_2$) | 0.6 | 0.5 |

(2.4)

All that may be shown is that, for each particle, the occurrence of both features forms a range described by (2.5), where $n(H_i)$ is the total number of exemplars $i$, $n(D_1|H_i)$ is the total number of $i$ with spin $\uparrow$, and $n(D_2|H_i)$ is the total number of $i$ with a positive charge,

$$n(D_1 \cap D_2|H_i) \in \begin{cases} \Big[n(D_1|H_i) + n(D_2|H_i) - n(H_i)\ , \ldots , \min(n(D_1|H_i), n(D_2|H_i))\Big] \\ \qquad \text{if } n(D_1|H_i) + n(D_2|H_i) > n(H_i)\ , \quad \text{or} \\ \Big[0\ , \ldots , \min(n(D_1|H_i), n(D_2|H_i))\Big] \\ \qquad \text{if } n(D_1|H_i) + n(D_2|H_i) \leq n(H_i)\ . \end{cases} \tag{2.5}$$

Specifically for (2.4) these ranges equate to

$$\begin{aligned} n(D_1 \cap D_2|H_1) &\in \{4,\ 5,\ 6\}\ , \\ n(D_1 \cap D_2|H_2) &\in \{2,\ 3,\ 4,\ 5\}\ . \end{aligned} \tag{2.6}$$

The simplest approach to resolving this problem is to naively ignore any intersection, or co-dependence, of the data and to directly multiply the marginal probabilities. Hence, given (2.4), the likelihood of particle $\alpha$ having the greatest occurrence of both spin $\uparrow$ and a positive charge would be calculated as

$$\begin{aligned} P(H_1|D_1 \cap D_2) &= \frac{0.5 \times 0.8 \times 0.6}{(0.5 \times 0.8 \times 0.6) + (0.5 \times 0.7 \times 0.5)} \\ &\approx 0.578\ . \end{aligned} \tag{2.7}$$

Yet, because the data intersect, this probability value is only one of a number which may be reasonably calculated. Alternatives include calculating a likelihood ratio using the mean value $\mu$ of the frequency ranges for each hypothesis

$$\begin{aligned} P(\mu[n(D_1 \cap D_2|H_1)]) &= \frac{1}{10} \times \frac{1}{3}(4 + 5 + 6) = 0.5\ , \\ P(\mu[n(D_1 \cap D_2|H_2)]) &= \frac{1}{10} \times \frac{1}{4}(2 + 3 + 4 + 5) = 0.35 \\ &\Rightarrow\ P(H_1|\mu D_1 \cap D_2) \approx 0.588\ ; \end{aligned} \tag{2.8}$$

and taking the mean value of the probability range derived from the frequency range

$$\begin{aligned} \min P(H_1|D_1 \cap D_2) &= \frac{4}{4+5}\ , \\ \max P(H_1|D_1 \cap D_2) &= \frac{6}{6+2} \\ \Rightarrow\ \mu[P(H_1|D_1 \cap D_2)] &\approx 0.597\ . \end{aligned} \tag{2.9}$$

Given this multiplicity of probability values, it would seem that none of these methods may lay claim to normativity. This problem of covariate overlap has, of course, been previously addressed within statistical literature. For instance, the "maximum likelihood" approach of Dempster, Laird, & Rubin has demonstrated how an "expectation-maximization" algorithm may be used to derive appropriate covariate overlap measures.[4] Indeed, the mathematical efficacy of this technique has been

confirmed by Wu.[6] However, it is difficult to see how such an iterative methodology can be employed without introducing axiomatic assumptions. Further, since any assumptions, irrespective of how benign they may appear, have the potential to skew results, what is required is an approach in which covariate overlaps can be automatically, and directly, calculated from contingency data.

## 3. A quantum mechanical proof of Bayes' theorem for independent data

Previously unconsidered, the quantum mechanical von Neumann axioms would seem to offer the most promise in this regard, since the re-conceptualization of covariate data as a quantum entangled system allows for statistical analysis with few, non-arbitrary assumptions. Unfortunately there are many conceptual difficulties that can arise here. For instance, a Dirac notation representation of (2.4) as a standard quantum superposition is

$$\begin{aligned}|\Psi\rangle = \frac{1}{\sqrt{N}}\Bigg[\alpha\Bigg(\sqrt{\frac{1}{3}}\,|4\rangle_{H1} + \sqrt{\frac{1}{3}}\,|5\rangle_{H1} + \sqrt{\frac{1}{3}}\,|6\rangle_{H1}\Bigg) \\ + \beta\Bigg(\sqrt{\frac{1}{4}}\,|2\rangle_{H2} + \sqrt{\frac{1}{4}}\,|3\rangle_{H2} + \sqrt{\frac{1}{4}}\,|4\rangle_{H2} + \sqrt{\frac{1}{4}}\,|5\rangle_{H2}\Bigg)\Bigg]\ . \end{aligned} \tag{3.10}$$

In this example, (3.10) cannot be solved since the possible values of $D_1 \cap D_2$ for each hypothesis (2.6) have been described as equal chance outcomes within a general superposition of $H_1$ and $H_2$, with the unknown coefficients $\alpha$ and $\beta$ assuming the role of the classical Bayesian likelihood ratio.

The development of an alternative quantum mechanical description necessitates a return to the simplest form of Bayes' theorem using the case of exclusive populations $H_i$ and data sets $D$, $\bar{D}$, such as given in (2.2). Here, the overall probability of $H_1$ may be simply calculated as

$$P(H_1) = \frac{n(H_1)}{n(H_1) + n(H_2)}\ . \tag{3.11}$$

The a priori uncertainty in (2.2) may be expressed by constructing a wave function in which the four data points are encoded as a linear superposition

$$\begin{aligned}|\Psi\rangle =& \alpha_{1,1}\,|H_1 \otimes D\rangle + \alpha_{1,2}\,\left|H_1 \otimes \bar{D}\right\rangle \\ &+ \alpha_{2,1}\,|H_2 \otimes D\rangle + \alpha_{2,2}\,\left|H_2 \otimes \bar{D}\right\rangle\ . \end{aligned} \tag{3.12}$$

Since there is no overlap between either $D$ and $\bar{D}$ or the populations $H_1$ and $H_2$, each datum automatically forms an eigenstate basis with the orthonormal conditions

$$\begin{aligned}\langle H_1 \otimes D|H_1 \otimes D\rangle = \left\langle H_1 \otimes \bar{D}|H_1 \otimes \bar{D}\right\rangle = 1 \\ \langle H_2 \otimes D|H_2 \otimes D\rangle = \left\langle H_2 \otimes \bar{D}|H_2 \otimes \bar{D}\right\rangle = 1 \\ \text{all other bra-kets } = 0\ , \end{aligned} \tag{3.13}$$

where the normalization of the wave function demands that

$$\langle\Psi|\Psi\rangle = 1 \ , \tag{3.14}$$

so that the sum of the modulus squares of the coefficients $\alpha_{i,j}$ gives a total probability of 1

$$|\alpha_{1,1}|^2 + |\alpha_{1,2}|^2 + |\alpha_{2,1}|^2 + |\alpha_{2,2}|^2 = 1 \ . \tag{3.15}$$

For simplicity let

$$\begin{aligned} x_1 = P(D|H_1), \ \ y_1 = P(\bar{D}|H_1) \ , \\ x_2 = P(D|H_2), \ \ y_2 = P(\bar{D}|H_2) \ , \\ X_1 = P(H_1), \ \ X_2 = P(H_2) \ . \end{aligned} \tag{3.16}$$

If the coefficients $\alpha_{i,j}$ from (3.12) are set as required by (2.2), it follows that

$$|\alpha_{1,1}|^2 = x_1, \ \ |\alpha_{1,2}|^2 = y_1, \ \ |\alpha_{2,1}|^2 = x_2, \ \ |\alpha_{2,2}|^2 = y_2 \ , \tag{3.17}$$

so that the normalised wave function $|\Psi\rangle$ is described by

$$\begin{aligned} |\Psi\rangle = \frac{1}{\sqrt{N}}(\sqrt{x_1}\,|H_1 \otimes D\rangle + \sqrt{y_1}\,\left|H_1 \otimes \bar{D}\right\rangle \\ + \sqrt{x_2}\,|H_2 \otimes D\rangle + \sqrt{y_2}\,\left|H_2 \otimes \bar{D}\right\rangle) \end{aligned} \tag{3.18}$$

for some normalization constant $N$.

The orthonormality condition (3.14) implies that

$$N = x_1 + y_1 + x_2 + y_2 = X_1 + X_2 \ , \tag{3.19}$$

thereby giving the full wave function description

$$|\Psi\rangle = \frac{\sqrt{x_1}\,|H_1 \otimes D\rangle + \sqrt{y_1}\,\left|H_1 \otimes \bar{D}\right\rangle + \sqrt{x_2}\,|H_2 \otimes D\rangle + \sqrt{y_2}\,\left|H_2 \otimes \bar{D}\right\rangle}{\sqrt{X_1 + X_2}} \ . \tag{3.20}$$

If the value of $P(H_1|D)$ is to be calculated, i.e., the property $D$ is observed, then the normalized wave function (3.12) necessarily collapses to

$$|\Psi'\rangle = \alpha_1\,|H_1 \otimes D_1\rangle + \alpha_2\,|H_2 \otimes D_1\rangle \ , \tag{3.21}$$

where the coefficients $\alpha_{1,2}$ may be determined by projecting $|\Psi\rangle$ on to the two terms in $|\Psi'\rangle$ using (3.13), giving

$$\begin{aligned} \alpha_1 = \langle\Psi'|H_1 \otimes D\rangle = \sqrt{\frac{x_1}{X_1 + X_2}} \ , \\ \alpha_2 = \langle\Psi'|H_2 \otimes D\rangle = \sqrt{\frac{x_2}{X_1 + X_2}} \ . \end{aligned} \tag{3.22}$$

Normalizing (3.21) with the coefficient $N'$

$$|\Psi'\rangle = \frac{1}{\sqrt{N'}}\Big(\sqrt{\frac{x_1}{X_1+X_2}}\,|H_1\otimes D\rangle + \sqrt{\frac{x_2}{X_1+X_2}}\,|H_2\otimes D\rangle\Big)\ , \tag{3.23}$$

and using the normalization condition (3.14), implies that

$$1 = \langle\Psi'|\Psi'\rangle = \frac{1}{N'}\Big(\frac{x_1}{X_1+X_2} + \frac{x_2}{X_1+X_2}\Big)$$

$$\rightarrow N' = \frac{x_1+x_2}{X_1+X_2}\ . \tag{3.24}$$

Thus, after collapse, the properly normalized wave function (3.23) becomes

$$|\Psi'\rangle = \sqrt{\frac{x_1}{x_1+x_2}}\,|H_1\otimes D\rangle + \sqrt{\frac{x_2}{x_1+x_2}}\,|H_2\otimes D\rangle\ \ , \tag{3.25}$$

which means that the probability of observing $|H_1\otimes D\rangle$ is

$$P(|H_1\otimes D\rangle) = \left(\sqrt{\frac{x_1}{x_1+x_2}}\right)^2 = \frac{\alpha_1^2}{\alpha_1^2+\alpha_2^2} = \frac{x_1}{x_1+x_2}\ . \tag{3.26}$$

This is entirely consistent with Bayes' theorem and demonstrates its derivation using quantum mechanical axioms.

## 4. Quantum likelihood ratios for co-dependent data

Having established the principle of using a quantum mechanical approach for the calculation of simple likelihood ratios with mutually exclusive data (2.2), it is now possible to consider the general case of $n$ hypotheses and $m$ data (4.27), where the data are co-dependent, or intersect.

| | $H_1$ | $H_2$ | $\cdots$ | $H_n$ |
|---|---|---|---|---|
| $D_1$ | $x_{1,1}$ | $x_{1,2}$ | $\cdots$ | $x_{1,n}$ |
| $D_2$ | $x_{2,1}$ | $x_{2,2}$ | $\cdots$ | $x_{2,n}$ |
| $\vdots$ | | $\vdots$ | | $\vdots$ |
| $D_m$ | $x_{m,1}$ | $x_{m,2}$ | $\cdots$ | $x_{m,n}$ |

(4.27)

Here the contingency table in (4.27) has been indexed using

$$x_{i,\alpha}\ , \qquad \alpha = 1,2,\ldots,n; \quad i = 1,2,\ldots,m\ . \tag{4.28}$$

While the general wave function remains the same as before, the overlapping data create non-orthonormal inner products which can be naturally defined as

$$\langle H_\alpha \otimes D_i | H_\beta \otimes D_j \rangle = c_{ij}^\alpha \delta_{\alpha\beta} \ , \ \ c_{ij}^\alpha = c_{ji}^\alpha \in \mathbb{R} \ , c_{ii}^\alpha = 1 \ . \tag{4.29}$$

Assuming, for simplicity, that the overlaps $c_{ij}^\alpha$ are real, then there is a symmetry in that $c_{ij}^\alpha = c_{ji}^\alpha$ for each $\alpha$. Further, for each $\alpha$ and $i$, the state is normalized, i.e., $c_{ii}^\alpha = 1$. The given independence of the hypotheses $H_\alpha$ also enforces the Kroenecker delta function, $\delta_{\alpha\beta}$.

The Hilbert space $V$ spanned by the kets $|H_\alpha \otimes D_i\rangle$ is $mn$-dimensional and, because of the independence of $H_\alpha$, naturally decomposes into the direct sum (4.30) with respect to the inner product, thereby demonstrating that the non-orthonormal conditions are the direct sum of $m$ vector spaces $V^\alpha$:

$$V = \mathrm{Span}(\{|H_\alpha \otimes D_i\rangle\}) = \bigoplus_{\alpha=1}^{n} V^\alpha \ , \qquad \dim V^\alpha = m \ . \tag{4.30}$$

Since the inner products are non-orthonormal, each $V^\alpha$ must be individually orthonormalised. Given that $V$ splits into a direct sum, this may be achieved for each subspace $V^\alpha$ by applying the Gram-Schmidt algorithm to $\{|H_\alpha \otimes D_i\rangle\}$ of $V$. Consequently, the orthonormal basis may be defined as

$$|K_i^\alpha\rangle = \sum_{k=1}^{n} A_{i,k}^\alpha \, |H_\alpha \otimes D_k\rangle \ , \qquad \left\langle K_i^\alpha | K_j^\alpha \right\rangle = \delta_{ij} \ , \tag{4.31}$$

for each $\alpha = 1, 2, \ldots, n$ with $m \times m$ matrices $A_{i,k}^\alpha$, for each $\alpha$.
Substituting the inner products (4.29) gives

$$\sum_{k,k'=1}^{m} A_{ik}^\alpha A_{jk'}^\alpha c_{kk'}^\alpha = \delta_{ij} \quad \forall \alpha = 1, 2, \ldots, n \ . \tag{4.32}$$

The wave-function may now be written as a linear combination of the orthonormalised kets $|K_i^\alpha\rangle$ with the coefficients $b_i^\alpha$, and may be expanded into the $|H_\alpha \otimes D_i\rangle$ basis using (4.31), i.e.,

$$|\Psi\rangle = \sum_{\alpha,i} b_i^\alpha \, |K_i^\alpha\rangle = \sum_{\alpha,i,k} b_i^\alpha A_{ik}^\alpha \, |H_\alpha \otimes D_k\rangle \ . \tag{4.33}$$

As with (3.17) from earlier, the coefficients in (4.33) should be set as required by the contingency table

$$\sum_i b_i^\alpha A_{i,k}^\alpha = \sqrt{x_{k\alpha}} \ , \tag{4.34}$$

where, to solve for the $b$-coefficients, (4.32) may be used to invert

$$\sum_{k,k'} \sum_i b_i^\alpha A_{ik}^\alpha A_{jk'} c_{kk'}^\alpha = \sum_{k,k'} \sqrt{x_{k\alpha}} A_{jk'}^\alpha c_{k'k}^\alpha \ , \tag{4.35}$$

giving

$$b_j^\alpha = \sum_{k,k'} \sqrt{x_{k\alpha}} A^\alpha_{jk'} c^\alpha_{kk'} \ . \tag{4.36}$$

Having relabelled the indices as necessary, a back-substitution of (4.34) into the expansion (4.33) gives

$$|\Psi\rangle = \sum_{\alpha,i,k} b_i^\alpha A^\alpha_{i,k} \, |H_\alpha \otimes D_k\rangle = \sum_{\alpha,k} \sqrt{x_{k\alpha}} \, |H_\alpha \otimes D_k\rangle \ , \tag{4.37}$$

which is the same as having simply assigned each ket's coefficient to the square root of its associated entry in the contingency table.

The normalization factor for $|\Psi\rangle$ is simply $1/\sqrt{N}$, where $N$ is the sum of the squares of the coefficients $b$ of the orthonormalised bases $|K_i^\alpha\rangle$,

$$\begin{aligned} N = \sum_{i,\alpha} (b_i^\alpha)^2 &= \sum_{i,\alpha} b_i^\alpha \left( \sum_{k,k'} \sqrt{x_{k\alpha}} A^\alpha_{k',i} c^\alpha_{kk'} \right) \\ &= \sum_{k,k',\alpha} \sqrt{x_{k\alpha} x_{k'\alpha}} c^\alpha_{kk'} \ . \end{aligned} \tag{4.38}$$

Thus, the final normalized wave function is

$$|\Psi\rangle = \frac{\sum\limits_{\alpha,k} \sqrt{x_{k\alpha}} \, |H_\alpha \otimes D_k\rangle}{\sqrt{\sum\limits_{i,j,\alpha} \sqrt{x_{i\alpha} x_{j\alpha}} c^\alpha_{ij}}} \ , \tag{4.39}$$

where $\alpha$ is summed from 1 to $n$, and $i, j$ are summed from 1 to $m$. Note that, in the denominator, the diagonal term $\sqrt{x_{i\alpha} x_{j\alpha}} c^\alpha_{ij}$, which occurs whenever $i = j$, simplifies to $x_{i\alpha}$ since $c^\alpha_{ii} = 1$ for all $\alpha$.

From (4.39) it follows that, exactly in parallel to the non-intersecting case, if all properties $D_i$ are observed simultaneously, the probability of any hypothesis $H_\alpha$, for a fixed $\alpha$, is

$$P(H_\alpha | D_1 \cap D_2 \ldots \cap D_m) = \frac{\sum\limits_i (b_i^\alpha)^2}{\sum\limits_{i,\beta} (b_i^\beta)^2} = \frac{\sum\limits_{i,j} \sqrt{x_{i\alpha} x_{j\alpha}} c^\alpha_{ij}}{\sum\limits_{i,j,\beta} \sqrt{x_{i\beta} x_{j\beta}} c^\beta_{ij}} \ . \tag{4.40}$$

In the case of non-even populations for each hypothesis (i.e., non-even priors), the calculated probabilities should be appropriately weighted.

## 5. Example solution

Returning to the problem presented in the contingency table (2.4), it is now possible to calculate the precise probability for a randomly selected particle with the properties of "spin ↑" and "charge +" being particle $\alpha$ ($H_1$). For this $2 \times 2$ matrix,

recalling from (4.29) that $c_{ii}^{\alpha} = 1$ and $c_{ij}^{\alpha} = c_{ji}^{\alpha}$, the general expression (4.40) may be written as

$$P(H_1|D_1 \cap D_2) = \frac{\sum_{i,j=1}^{2} \sqrt{x_{i,1}x_{j,1}}c_{ij}^{1}}{\sum_{i,j=1}^{2}\sum_{\alpha=1}^{2} \sqrt{x_{i\alpha}x_{j\alpha}}c_{ij}^{\alpha}}$$

$$= \frac{\sqrt{x_{1,1}^2}c_{1,1}^{1} + \sqrt{x_{2,1}^2}c_{2,2}^{1} + \sqrt{x_{1,1}x_{2,1}}c_{1,2}^{1} + \sqrt{x_{2,1}x_{1,1}}c_{2,1}^{1}}{\sum_{\alpha=1}^{2} \sqrt{x_{1,\alpha}^2}c_{1,1}^{1} + \sqrt{x_{2,\alpha}^2}c_{2,2}^{1} + \sqrt{x_{1,\alpha}x_{2,\alpha}}c_{1,2}^{1} + \sqrt{x_{2,\alpha}x_{1,\alpha}}c_{2,1}^{1}}$$

$$= \frac{x_1 + y_1 + 2c_1\sqrt{x_1y_1}}{x_1 + x_2 + y_1 + y_2 + 2c_1\sqrt{x_1y_1} + 2c_2\sqrt{x_2y_2}} \ , \tag{5.41}$$

where, adhering to the earlier notation (3.16),

$$\begin{aligned} x_1 = x_{1,1} = P(D_1|H_1), \ \ y_1 = x_{2,1} = P(D_2|H_1) \ , \\ x_2 = x_{1,2} = P(D_1|H_2), \ \ y_2 = x_{2,2} = P(D_2|H_2) \ , \\ X_1 = P(H_1), \ \ X_2 = P(H_2) \ , \end{aligned} \tag{5.42}$$

and, for brevity, $c_1 := c_{1,2}^{1}$ , $c_2 := c_{1,2}^{2}$ . For simplicity, $P(H_i|D_1 \cap D_2)$ will henceforth be denoted as $P_i$. Implementing (5.41) is dependent upon deriving solutions for the yet unknown expressions $c_i$, $i = 1, 2$ which govern the extent of the intersection in (4.29). This can only be achieved by imposing reasonable constraints upon $c_i$ which have been inferred from expected behaviour and known outcomes, i.e., through the use of boundary values and symmetries. Specifically, these constraints are:

**Data dependence.** The expressions $c_i$ must, in some way, be dependent upon the data given in the contingency table, i.e.,

$$\begin{aligned} c_1 = c_1(x_1, y_1, x_2, y_2; X_1, X_2) \ , \\ c_2 = c_2(x_1, y_1, x_2, y_2; X_1, X_2) \ . \end{aligned} \tag{5.43}$$

**Probability.** The calculated values for $P_i$ must fall between 0 and 1. Since $x_i$ and $y_i$ are positive, it suffices to take

$$-1 \ < \ c_i(x_1, \ y_1, \ x_2, \ y_2) \ < \ 1 \ . \tag{5.44}$$

**Complementarity.** The law of total probability dictates that

$$P_1 + P_2 = 1 \ , \tag{5.45}$$

which can easily be seen to hold.

**Symmetry.** The exchanging of rows within the contingency tables should not affect the calculation of $P_i$. In other words, for each $i = 1, 2$, $P_i$ is invariant under $x_i \leftrightarrow y_i$. This constraint implies that

$$c_i(x_1, y_1, x_2, y_2) = c_i(y_1, x_1, y_2, x_2) \ . \tag{5.46}$$

Equally, if the columns are exchanged then $P_i$ must map to each other, i.e., for each $i = 1, 2$ then $P_1 \leftrightarrow P_2$ under $x_1 \leftrightarrow x_2$, $y_1 \leftrightarrow y_2$ which gives the further constraint that

$$c_1(x_1, y_1, x_2, y_2) = c_2(x_2, y_2, x_1, y_1) \ . \tag{5.47}$$

**Known values.** There are a number of contingency table structures which give rise to a known probability, i.e.,

| | $H_1$ | $H_2$ |
|---|---|---|
| $D_1$ | 1 | 1 |
| $D_2$ | $m$ | $n$ |

$$\rightarrow \qquad P_1 = \frac{m}{m+n}$$

| | $H_1$ | $H_2$ |
|---|---|---|
| $D_1$ | $m$ | $n$ |
| $D_2$ | 1 | 1 |

$$\rightarrow \qquad P_1 = \frac{m}{m+n}$$

| | $H_1$ | $H_2$ |
|---|---|---|
| $D_1$ | $n$ | $m$ |
| $D_2$ | $m$ | $n$ |

$$\rightarrow \qquad P_1 = \frac{1}{2}$$

| | $H_1$ | $H_2$ |
|---|---|---|
| $D_1$ | $n$ | $n$ |
| $D_2$ | $m$ | $m$ |

$$\rightarrow \qquad P_1 = \frac{1}{2}$$

| | $H_1$ | $H_2$ |
|---|---|---|
| $D_1$ | $m$ | $m$ |
| $D_2$ | $m$ | $m$ |

$$\rightarrow \qquad P_1 = \frac{1}{2} \ , \tag{5.48}$$

where $m, n$ are positively valued probabilities. For such contingency tables the correct probabilities should always be returned by $c_i$. Applying this principle to (5.41) gives the constraints

$$\frac{m}{m+n} = \frac{2c_1(m,1,n,1)\sqrt{m}+m+1}{2c_1(m,1,n,1)\sqrt{m}+2c_2(m,1,n,1)\sqrt{n}+m+n+2} , \quad (5.49)$$

$$\frac{1}{2} = \frac{2c_1(n,m,m,n)\sqrt{m}\sqrt{n}+m+n}{2c_1(n,m,m,n)\sqrt{m}\sqrt{n}+2c_2(n,m,m,n)\sqrt{m}\sqrt{n}+2m+2n} , \quad (5.50)$$

$$\frac{1}{2} = \frac{2c_1(n,m,n,m)\sqrt{m}\sqrt{n}+m+n}{2c_1(n,m,n,m)\sqrt{m}\sqrt{n}+2c_2(n,m,n,m)\sqrt{m}\sqrt{n}+2m+2n} . \quad (5.51)$$

**Non-homogeneity.** Bayes' theorem returns the same probability for any linearly scaled contingency tables, e.g.,

$$x_1 \to 1.0,\ y_1 \to 1.0,\ x_2 \to 1.0,\ y_2 \to 0.50 \Rightarrow P_1 \approx 0.667 , \quad (5.52)$$
$$x_1 \to 0.5,\ y_1 \to 0.5,\ x_2 \to 0.5,\ y_2 \to 0.25 \Rightarrow P_1 \approx 0.667 . \quad (5.53)$$

While homogeneity may be justified for conditionally independent data, this is not the case for intersecting, co-dependent data since the act of scaling changes the nature of the intersections and the relationship between them. This may be easily demonstrated by taking the possible value ranges for (5.52) and (5.53), calculated using (2.5), which are

$$\begin{aligned} \text{Eq. (5.52)} \Rightarrow & (D_1 \cap D_2)|H_1 = \{1\} , \\ & (D_1 \cap D_2)|H_2 = \{0.5\} , \end{aligned}$$

$$\begin{aligned} \text{Eq. (5.53)} \Rightarrow & (D_1 \cap D_2)|H_1 = \{0.0 \ldots 0.5\} , \\ & (D_1 \cap D_2)|H_2 = \{0.0 \ldots 0.25\} . \end{aligned} \quad (5.54)$$

The effect of scaling has not only introduced uncertainty where previously there had been none, but has also introduced the possibility of 0 as a valid answer for both hypotheses. Further, the spatial distance between the hypotheses has also decreased. For these reasons it would seem unreasonable to assert that (5.52) and (5.53) share the same likelihood ratio.

Using these principles and constraints it becomes possible to solve $c_i$. From the principle of symmetry it follows that

$$\begin{aligned} & c_1(n,m,m,n) = c_2(m,n,n,m) = c_2(n,m,m,n) , \\ & c_1(n,m,n,m) = c_2(n,m,n,m) = c_2(n,m,n,m) , \end{aligned} \quad (5.55)$$

and that the equalities (5.50), (5.51) for $P_i = 0.5$ automatically hold. Further, (5.49) solves to give

$$c_2(m,1,n,1) = \frac{2\sqrt{mn}c_1(m,1,n,1) - m + n}{2m\sqrt{n}} \ , \tag{5.56}$$

which, because $c_1(n,1,m,1) = c_2(m,1,n,1)$, finally gives

$$c_1(n,1,m,1) = \frac{2\sqrt{mn}c_1(m,1,n,1) - m + n}{2m\sqrt{n}} \ . \tag{5.57}$$

Substituting $g(m,n) := \sqrt{n}c_1(m,1,n,1)$ transforms (5.57) into an anti-symmetric bivariate functional equation in $m,n$,

$$g(m,n) - g(n,m) = \frac{m}{2\sqrt{mn}} - \frac{n}{2\sqrt{mn}} \ , \tag{5.58}$$

whose solution is $g(m,n) = \frac{m}{2\sqrt{mn}}$ .

This gives a final solution for the coefficients $c_{1,2}$ of

$$\begin{aligned} c_1(x_1,y_1,x_2,y_2) &= \frac{\sqrt{x_1y_1}}{2x_2y_2} \ , \\ c_2(x_1,y_1,x_2,y_2) &= \frac{\sqrt{x_2y_2}}{2x_1y_1} \ . \end{aligned} \tag{5.59}$$

Thus, substituting (5.59) into (5.41) gives the likelihood ratio expression of,

$$P(H_1|D_1 \cap D_2) = \frac{\frac{x_1y_1}{x_2y_2} + x_1 + y_1}{\frac{x_1y_1}{x_2y_2} + x_1 + y_1 + \frac{x_2y_2}{x_1y_1} + x_2 + y_2} \ . \tag{5.60}$$

Given that the population sizes of $H_1$ and $H_2$ are the same, no weighting needs to take place. Hence, the value of $P(H_1|D_1 \cap D_2)$ for (2.4) may now be calculated to be

$$P(H_1|D_1 \cap D_2) \approx 0.5896 \ . \tag{5.61}$$

## 6. Discussion

One of the greatest obstacles in developing any statistical approach is demonstrating correctness. This formula is no different in that respect. If correctness could be demonstrated then, a priori, there would be an appropriate existing method which would negate the need for a new one. All that may be hoped for in any approach is that it generates appropriate answers when they are known, reasonable answers for all other cases, and that these answers follow logically from the underlying mathematics.

However, what is clear is that the limitations of the naive Bayes' classifier render any calculations derived from it open to an unknown margin of error. Given the

importance of accurately deriving likelihood ratios this is troubling. This is especially true when the statistical tolerance of calculations is marginal.

As a quantum mechanical methodology this result is able to calculate accurate, iteration free, likelihood ratios which fall beyond the scope of existing statistical techniques, and offers a new theoretical approach within both statistics and physics. Further, through the addition of a Hamiltonian operator to introduce time-evolution, it can offer likelihood ratios for future system states with appropriate updating of the contingency table. In contrast, Bayes' theorem is unable to distinguish directly between time-dependent and time-independent systems. This may lead to situations where the process of contingency table updating results in the same decisions being made repeatedly with the appearance of an ever increasing degree of certainty. Indeed, from (3.26), it would seem that the naive Bayes' classifier is only a special case of a more complex quantum mechanical framework, and may only be used where the exclusivity of data is guaranteed.

The introduction of a Hamiltonian operator, and a full quantum dynamical formalism, is in progress, and should have profound implications for the physical sciences. Inevitably, such a formalism will require a sensible continuous classical limit. In other words, the final expressions for the likelihood ratios should contain a parameter, in some form of $\hbar$, which, when going to 0, reproduces a classically known result. For example, the solutions to (5.59) could be moderated as

$$c_1(x_1, y_1, x_2, y_2) = \frac{\sqrt{x_1 y_1}}{2x_2 y_2}(1 - \exp(-\hbar)) \ ,$$

$$c_2(x_1, y_1, x_2, y_2) = \frac{\sqrt{x_2 y_2}}{2x_1 y_1}(1 - \exp(-\hbar)) \ , \tag{6.62}$$

so that in the limit of $\hbar \to 0$, the intersection parameters, $c_1$ and $c_2$, vanish to return the formalism to the classical situation of independent data.

## 7. Conclusion

This article has demonstrated both theoretically, and practically, that a quantum mechanical methodology can overcome the axiomatic limitations of classical statistics. In doing so, it challenges the orthodoxy of de Finetti's epistemological approach to statistics by demonstrating that it is possible to derive "real" likelihood ratios from information systems without recourse to arbitrary and subjective evaluations.

While further theoretical development work needs to be undertaken, particularly with regards to the application of these mathematics in other domains, it is hoped that this article will help advance the debate over the nature and meaning of statistics within the physical sciences.

**Acknowledgments**

YHH would like to thank the Science and Technology Facilities Council, UK, for grant ST/J00037X/1, the Chinese Ministry of Education, for a Chang-Jiang Chair Professorship at NanKai University as well as the City of Tian-Jin for a Qian-Ren Scholarship, and Merton College, Oxford, for her enduring support.

**References**

1. de Finetti, B., *Theory of probability: A critical introductory treatment* (2 vols.), trans. Machí, A., and Smith, A. (Wiley, New York, 1974).
2. Caves, C., Fuchs, C., and Schack, R., Conditions for compatibility of quantum-state assignments. *Physical Review A* **66(6)** (2002) 062111(1)–062111(11). doi: 10.1103/PhysRevA.66.062111
3. Timpson, C. G., Quantum Bayesianism: A study. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **39(3)** (2008) 579–609. doi: 10.1016/j.shpsb.2008.03.006
4. Dempster, A. P., Laird, N. M., and Rubin, D. B., Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **39(1)** (1977) 1–38.
5. Oaksford, M. and Chater, N., *Bayesian rationality. The probabilistic approach to human reasoning* (Oxford University Press, Oxford, 2007).
6. Wu, C. F. J., On the convergence properties of the EM algorithm. *The Annals of Statistics* **11(1)** (1983) 95–103. doi: 10.1214/aos/1176346060